

覃卫坚,何莉阳,蔡悦幸.广西暴雨集中度智能气候预测方法研究[J].气象研究与应用,2024,45(3):12-20.

QIN Weijian, HE Liyang, CAI Yuexing. Research on intelligent climate prediction methods of rainstorm concentration in Guangxi [J]. Journal of Meteorological Research and Application, 2024, 45(3): 12-20.

## 广西暴雨集中度智能气候预测方法研究

覃卫坚, 何莉阳, 蔡悦幸

(广西壮族自治区气候中心, 南宁 530022)

**摘要:** 利用1961—2023年广西79个气象观测站逐日降水和国家气候中心大气环流、海温指数资料,构建广西暴雨集中度计算方法,基于逐步回归方法、粒子群-神经网络、随机森林算法,建立暴雨集中度气候预测模型。结果表明,广西存在以桂林和柳州两市北部为中心的桂东北地区、以“东兰、巴马、凤山”为中心的桂西山区和沿海地区三个暴雨集中度高值区,暴雨集中度异常大小基本反映发生洪涝和干旱灾害的严重程度。经过2020—2023年气候预测试验,粒子群-神经网络算法预测效果最好,其次为随机森林算法,第三是逐步回归方法。

**关键词:** 暴雨;集中度;集中期;粒子群-神经网络;随机森林算法

**中图分类号:** P456

**文献标识码:** A

**doi:** 10.19849/j.cnki.CN45-1356/P.2024.3.02

在年度气候趋势预测中,人们除了关注气温、降水气候趋势预测外,更加关注当年是否出现旱涝灾害,而暴雨集中期和集中度指标,正好可以大致反映暴雨发生的时间、暴雨集中程度等旱涝分布情况,通常暴雨集中度越大,洪涝灾害可能越严重,而集中度越小,则可能发生干旱。如何表达降水年内逐日变化的集中程度,一些专家把一年的时间按顺序排列成一个圆周,把降水事件发生的时间当成矢量方向,降水事件作为模,用正弦和余弦三角函数分别对矢量方向的角度进行计算,分别乘以模,然后相加,最后再除以降水量(或暴雨降水量,或极端降水量累加值),即得到的值用来表征降水年内非均匀分布程度<sup>[1-5]</sup>。假设降水发生事件分布均匀,那么计算得到的矢量累计值等于零,表示降水在年内分布很均匀,否则不均匀,矢量累计值越大、不均匀程度越大。这种计算方法让集中度既能反映暴雨集中程度,又能体现出洪涝灾害程度,这在以往研究中常常被忽视。

海洋和大气因子往往会影响年内暴雨气候异常,如Karori等<sup>[6]</sup>指出厄尔尼诺中部型对华南夏季

降水影响显著;Zhu等<sup>[7]</sup>认为热带太平洋海温西暖东冷模态使海洋性大陆地区对流旺盛,通过哈德莱环流使华南地区春季降水偏少,夏季热带印度洋暖海温使海洋性大陆地区对流受到抑制,通过哈德莱环流有利于东亚季风槽偏强,导致华南地区降水偏多;Nguyen-Le等<sup>[8]</sup>发现La Nina年中南半岛夏季雨季开始期偏早;章开美等<sup>[9]</sup>研究得到前汛期类La Nina型时,有利于华南前汛期发生极端强降水,而类El Nino型时,则不利于华南前汛期极端强降水发生;王彦明等<sup>[10]</sup>研究指出PDO冷位相年,华南前汛期发生典型持续性暴雨过程的概率比暖位相大,且暴雨强度偏强,持续时间长;覃卫坚等<sup>[11]</sup>研究发现PDO冷位相年广西大范围暴雨日数偏多;蔡学湛<sup>[12]</sup>研究发现前冬春青藏高原雪盖多(少)雪年有利于华南前汛期雨涝(干旱);Wang等<sup>[13]</sup>研究发现青藏高原南部积雪偏多不利于华南夏季降水,而青藏高原西部和北部积雪偏多分别有利于中国华南和东南部地区夏季降水发生。可见,可以通过前期海温等信号预测年暴雨气候异常情况。

机器学习方法具有较强的处理非线性问题的

收稿日期: 2024-05-15

基金项目: 广西科技计划项目(桂科AB21075005)、广西气象科技计划项目(2023Z005)

第一作者: 覃卫坚(1971-), 博士, 正高级工程师, 主要从事气候变化和气候预测研究。E-mail: qinweijian2008@126.com

能力,在气象预报中有很好的应用效果,如陆虹等<sup>[14]</sup>把粒子群-神经网络方法应用在广西冷湿天气预报中,预报效果明显;覃卫坚等<sup>[15-16]</sup>使用机器学习预测方法提高寒露风气候预测准确率;孔庆燕等<sup>[17]</sup>、吴建生等<sup>[18]</sup>使用支持向量机等机器学习算法提高降水量预报效果;田心如等<sup>[19]</sup>把机器学习算法在夏季空调负荷预报中应用,预报准确度有明显提高;Kim and Kim<sup>[20]</sup>把随机森林方法应用于城市洪水灾害等级预测中,提高预测准确率。可见,机器学习算法在各种气象预报中取得很好的应用效果,在广西暴雨集中度气候预测中应用值得尝试。本研究拟构建暴雨集中度、集中期指标计算方法,建立基于机器学习算法的气候预测模型,尝试开展暴雨非均匀性分布的气候预测,为气候预测服务提供参考。

## 1 资料与方法

### 1.1 资料

使用1961—2023年广西79个国家气象观测站逐日降水资料、国家气候中心整理的88项环流特征量和26项海温指数。

### 1.2 暴雨集中度计算方法

暴雨集中度(RCD)计算公式为:

$$RCD_i = \sqrt{R_{xi}^2 + R_{yi}^2} / R_i \quad (1)$$

暴雨集中期(RCP)计算公式为:

$$RCP_i = \arctan(R_{xi}/R_{yi}) \quad (2)$$

其中

$$R_{xi} = \sum_{j=1}^N r_{ij} \times \sin \theta_j \quad (3)$$

$$R_{yi} = \sum_{j=1}^N r_{ij} \times \cos \theta_j \quad (4)$$

其中 $x, y$ 指矢量分解后相互垂直的两个方向, $r_{ij}$ 某日20:00至次日20:00降水量 $\geq 50$  mm的降水量, $R_i$ 为研究时段中雨量级以上总降水量, $\theta_i$ 为研究时段内对应的方位角, $i$ 为年份, $j$ 为研究时段内的日序。

在模和降水量累加值的选择方面考虑降水的量级,在矢量方向选择上,使用暴雨发生日对应的角度。在模和降水累加值选择上,由于小雨对洪涝灾害的影响小,因此使用中雨以上的总降水量作为模,计算得到的集中度和洪涝灾害面积的相关系数达到0.5左右为最大,而使用大雨、暴雨量级以上总降水量计算得到的相关系数略低。因此(1)式采用研究时段中雨量级以上的总降水量,这样计算得到

的集中度与暴雨灾害严重程度关系更加密切。

### 1.3 粒子群-神经网络方法

粒子群-神经网络计算<sup>[14, 21-22]</sup>步骤为:

第一步:输入训练样本,计算每个粒子的适应值

$$F(w, v, \theta, \gamma) = \frac{1}{1 + \min E(w, v, \theta, \gamma)} \quad (5)$$

并且初始化个体经历最好位置 $P_{best}(f)$ ,以及群体经历的最好位置 $P_{gbest}(f)$ 。

第二步:计算调校粒子的适应值与所经历的最好位置的适应值达到较好,则将其作为每个粒子当前的最好位置;同样,计算调校其适应值与全局所经历的最好位置的适应值到较好,则将其作为当前的全局最优位置。

第三步:粒子的速度进化方程为:

$$v_{ij}(f+1) = \omega \cdot v_{ij}(f) + c_1 r_1 (P_{best}(f) - x_{ij}(f)) + c_2 r_2 (P_{gbest}(f) - x_{ij}(f)) \quad (6)$$

$$\omega(t) = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{iter_{max}} \times I \quad (7)$$

其中 $\omega_{max}$ 、 $\omega_{min}$ 分别是惯性权重的最大值和最小值, $I, iter_{max}$ 分别是当前迭代次数和最大的迭代次数, $c_1, c_2$ 为记忆因子, $r_1, r_2$ 为(0,1)间的随机数。

第四步:位置权重系数进化方程<sup>[23-24]</sup>:

$$x_{ij}(t+1) = x(t) + v_{ij}(t+1) \quad (8)$$

第五步:位置矩阵中的结构矩阵 $S$ 的进化方程为<sup>[25]</sup>:

$$S_{ij}(t+1) = \begin{cases} 0, & r \geq \frac{1}{1 + \exp(-v_{ij}(t+1))} \\ 1, & r < \frac{1}{1 + \exp(-v_{ij}(t+1))} \end{cases} \quad (9)$$

其中 $r$ ,为[0,1]均匀分布的随机数。

反复第一步到第五步计算,优化粒子速度和位置,直到适应值满足或达到总的进化代数。

### 1.4 随机森林算法

使用多棵决策树对样本进行训练和预测的分类器,由不完全相同的单棵决策树组成,利用多棵决策树投票机制来决定最终的分类,这是随机森林算法。设定含有 $N$ 个样本的原始样本集,从原始样本集中随机抽样,组成多个训练集,建立 $N$ 棵决策树<sup>[26-27]</sup>:

$$\{h(x, \theta_n), n = 1, 2, \dots, N\} \quad (10)$$

为输入的自变量和因变量,为服从独立同分布随机向量。

在训练决策树模型的节点时,随机从节点上所有样本特征中选择一部分样本特征,以其中最优的一个特征来划分决策树的左右子树,训练结束后进行投票得到所有模态的平均值作为输出:

$$h(x) = \frac{1}{N} \sum_{n=1}^N h(x, \theta_n) \quad (11)$$

## 2 暴雨集中度气候特征

利用 1961—2023 年全区逐日降水资料统计得到广西暴雨集中度,全区平均暴雨集中度为 0.25,呈现出桂北及沿海地区大于桂中地区的分布特征(图 1a)。广西存在三个暴雨集中度高值区:一是以桂林和柳州二市北部为中心的桂东北地区,二是以“东兰、巴马、凤山”为中心的桂西山区,三是沿海地区。这三个暴雨高值区集中度均在 0.25 以上,其中位于沿海地区的东兴市暴雨集中度最大,达到 0.42,其次是北海达到 0.41,可见沿海地区暴雨发生比较集中,容易形成暴雨洪涝和次生灾害,为暴雨洪涝灾害高发区。

计算广西逐年暴雨集中度,得到历年变化曲线图(图 1b),从一元回归方程可看出,广西暴雨集中度总体呈逐年增大的趋势,总体趋势变化不显著,但年际变化波动大,最大值是最小值的两倍。1994 和 1998 年暴雨集中度最大,分别高达 0.31,广西均发生严重的暴雨洪涝灾害。暴雨集中度最小的年

份为 1963 年,当年暴雨集中度为 0.16,广西发生严重的干旱灾害。可见暴雨集中度异常年份对应着严重的旱涝灾害。

## 3 暴雨集中度气候预测方法

### 3.1 计算暴雨集中度预测因子

对 1961—2019 年广西 79 站暴雨集中度进行 EOF 分解,得到主要空间模态和各模态的时间系数。各特征向量能够反映出暴雨集中度变化的空间结构,第一模态是暴雨集中度变化最具有代表性的分布场,其次为第二模态、第三模态等,前五个模态的方差贡献率分别为 16.7%、9.8%、4.9%、4.7%、4.5%,为了减少计算量,只计算前五个模态的时间系数。

计算暴雨集中度前五个模态时间系数与前一年大气、海温指数资料的相关系数,相关显著的指数进入因子集。利用逐步回归方法对因子进行计算,建立逐步回归方程。

第一模态时间系数逐步回归预报方程:

$$y = -1.527 + 0.001x_1 + 0.082x_3 + 0.072x_5 + 0.034x_8 - 0.051x_{12} + 0.068x_{16} \quad (12)$$

( $F=2$ ,  $\sigma=0.287$ ,  $R=0.657$ )

第二模态时间系数逐步回归预报方程:

$$y = -0.307 - 0.072x_2 + 0.033x_{13} - 0.040x_{16} - 0.004x_{24} + 0.442x_{25} - 0.130x_{27} - 0.150x_{39} \quad (13)$$

( $F=2$ ,  $\sigma=0.183$ ,  $R=0.785$ )

第三模态时间系数逐步回归预报方程:

$$y = 6.25 - 0.003x_8 - 0.05x_{10} - 0.038x_{11} + 0.008x_{15} - 0.001x_{16} + 0.021x_{37} + 0.066x_{42} \quad (14)$$

( $F=2$ ,  $\sigma=0.132$ ,  $R=0.774$ )

第四模态时间系数逐步回归预报方程:

$$y = -1.22959 + 0.00010x_9 + 0.00625x_{11} + 0.03296x_{12} - 0.00105x_{14} + 0.00513x_{15} - 0.00271x_{16} + 0.11176x_{22} + 0.01731x_{28} - 0.20522x_{34} + 0.16184x_{36} - 0.06126x_{45} \quad (15)$$

( $F=2$ ,  $\sigma=0.112$ ,  $R=0.851$ )

第五模态时间系数逐步回归预报方程:

$$y = 3.75649 + 0.17734x_1 - 0.01512x_4 + 0.03345x_9 - 0.00052x_{12} - 0.00041x_{13} - 0.00052x_{14} - 0.03989x_{15} - 0.00546x_{16} - 0.00511x_{17} - 0.00833x_{22} - 0.02061x_{26} \quad (16)$$

( $F=2$ ,  $\sigma=0.116$ ,  $R=0.83$ )

通过逐步回归方法再次筛选,得到第一模态 6 个因子,第二、三模态各 7 个因子,第四、五模态各 11 个因子,具体因子如表 1 所示。

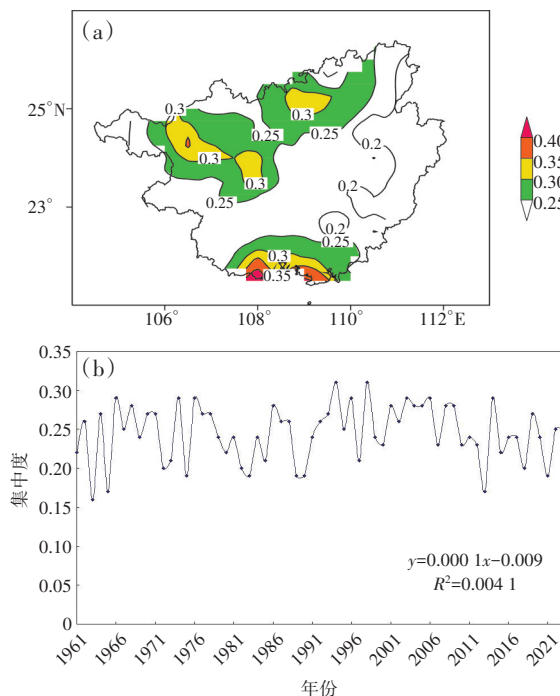


图1 广西暴雨集中度空间分布(a)和年际变化图(b)

表 1 经逐步回归方法筛选得到的预报因子

模态	因子序号	因子名称
第一模态	X <sub>1</sub>	9月北半球极涡中心经向位置指数
	X <sub>3</sub>	8月欧亚纬向环流指数
	X <sub>5</sub>	2月亚洲经向环流指数
	X <sub>8</sub>	5月赤道中东太平洋 200 hPa 纬向风指数
	X <sub>12</sub>	5月 850 hPa 中太平洋信风指数
	X <sub>16</sub>	9月 850 hPa 东太平洋信风指数
第二模态	X <sub>2</sub>	5月东太平洋副高面积指数
	X <sub>13</sub>	6月大西洋欧洲区极涡面积指数
	X <sub>16</sub>	6月欧亚经向环流指数
	X <sub>24</sub>	4月南极涛动指数
	X <sub>25</sub>	3月太平洋-北美遥相关型指数
	X <sub>27</sub>	2月极地-欧亚遥相关型指数
	X <sub>39</sub>	1月副热带南印度洋偶极子指数
第三模态	X <sub>8</sub>	3月南海副高强度指数
	X <sub>10</sub>	12月北美区极涡面积指数
	X <sub>11</sub>	11月大西洋欧洲区极涡面积指数
	X <sub>15</sub>	8月北半球极涡中心纬向位置指数
	X <sub>16</sub>	8月北半球极涡中心强度指数
	X <sub>37</sub>	7月西太平洋暖池面积指数
	X <sub>42</sub>	11月类 ENSO 指数
第四模态	X <sub>9</sub>	5月北大西洋-欧洲区极涡强度指数
	X <sub>11</sub>	5月北半球极涡中心纬向位置指数
	X <sub>12</sub>	1月亚洲经向环流指数
	X <sub>14</sub>	2月印槽强度指数
	X <sub>15</sub>	6月北极涛动指数
	X <sub>16</sub>	9月南极涛动指数
	X <sub>22</sub>	1月北太平洋遥相关型指数
	X <sub>28</sub>	5月北大西洋-欧洲环流 W 型指数
	X <sub>34</sub>	5月热带南大西洋海温指数
	X <sub>36</sub>	7月热带南大西洋海温指数
第五模态	X <sub>45</sub>	12月副热带南印度洋偶极子指数
	X <sub>1</sub>	1月北大西洋副高面积指数
	X <sub>4</sub>	10月西太平洋副高脊线位置指数
	X <sub>9</sub>	5月北美区极涡面积指数
	X <sub>12</sub>	1月北半球极涡中心经向位置指数
	X <sub>13</sub>	12月北半球极涡中心经向位置指数
	X <sub>14</sub>	9月北半球极涡中心强度指数
	X <sub>15</sub>	5月欧亚经向环流指数
	X <sub>16</sub>	1月东亚槽位置指数
	X <sub>17</sub>	6月北极涛动指数
	X <sub>22</sub>	6月赤道中东太平洋 200 hPa 纬向风指数
	X <sub>26</sub>	6月 850 hPa 西太平洋信风指数

3.2 基于逐步回归方法的暴雨集中度预测和检验

利用暴雨集中度前五个模态时间系数与前一年大气、海温指数的相关显著的因子集,使用逐步回归方法建立预报方程,对 2020—2023 年逐年暴雨集中度距平值进行预测试验,结果和实况距平值对

比如图 2 所示。从预报结果来看,逐步回归方法预测基本上能够把异常偏大、偏小区域预报出来。4 a 预报与实况距平对比符号相同率和均方根误差 (*RMSE*)如表 2 所示,同号率平均为 64.557%,均方根误差平均为 0.101(表 2)。



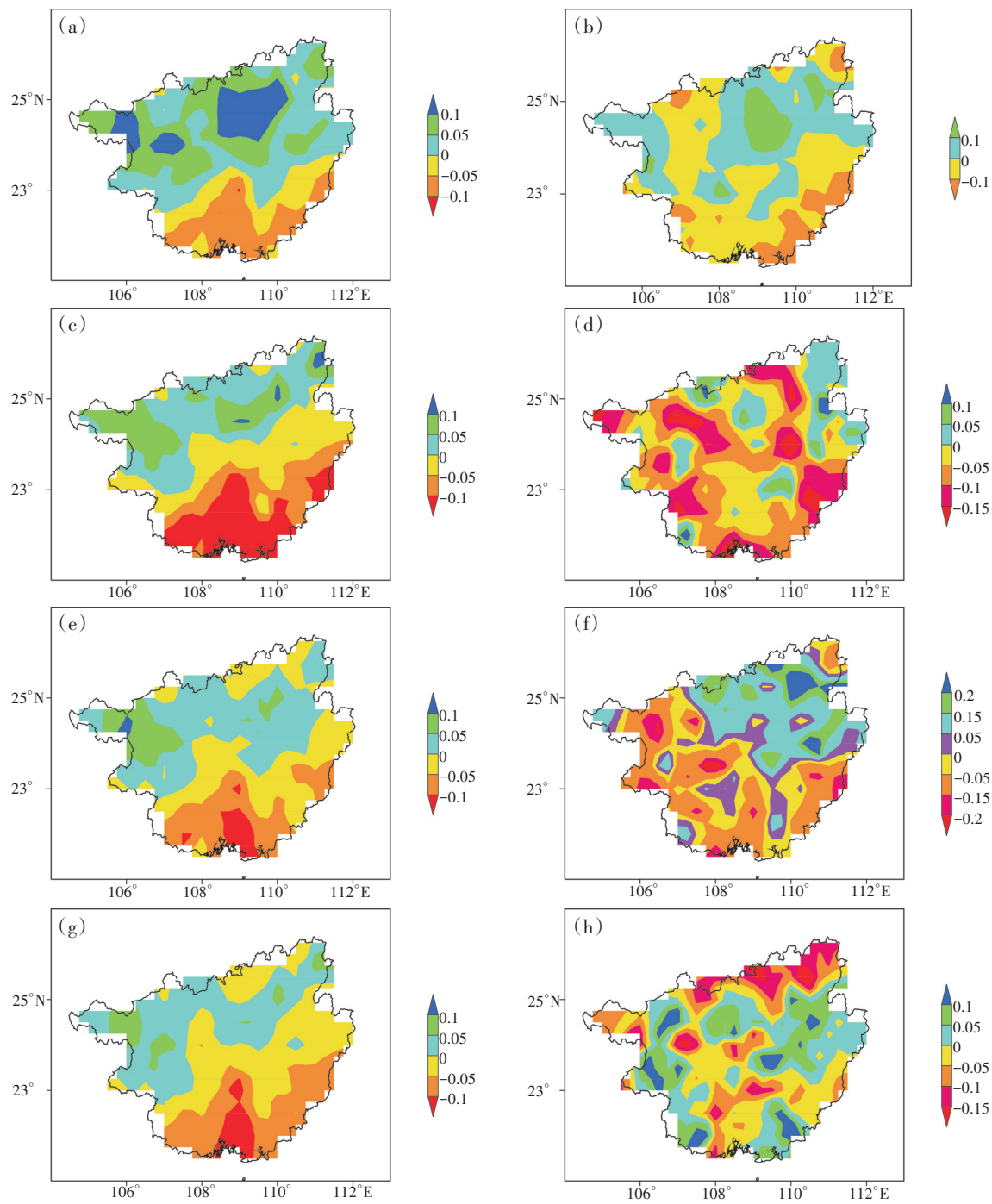


图2 逐步回归方法暴雨集中度气候预测图(a、c、e、g)和实况距平图(b、d、f、h)  
(a-b)2020年;(c-d)2021年;(e-f)2022年;(g-h)2023年

表2 各方法预测和实况距平的相同符号率及 *RMSE*

	逐步回归方法		粒子群-神经网络		随机森林算法	
	同号率/%	<i>RMSE</i>	同号率/%	<i>RMSE</i>	同号率/%	<i>RMSE</i>
2020	73.418	0.086	73.418	0.082	77.215	0.085
2021	67.089	0.102	82.278	0.079	82.278	0.083
2022	62.025	0.116	60.759	0.113	56.962	0.116
2023	55.696	0.099	68.354	0.088	67.089	0.087
平均	64.557	0.101	71.202	0.091	70.886	0.093

### 3.3 基于粒子群-神经网络算法的暴雨集中度预测和检验

通过逐步回归方程筛选得到预测因子,使用粒子群-神经网络算法建立预报模型。粒子群-神经网络预报模型输出节点个数为1,隐节点下限为0.3,隐节点上限为1.5,目标误差为0.01,学习速率为0.5,动量因子为0.75,训练次数为200,个体最优导向系数为2,全局最优导向系数为2,粒子位置下限为-3,粒子位置上限为3,种群规模为50,最大迭代次数为100。

对2020—2023年逐年暴雨集中度距平值进行预测试验,预测结果和实况距平对比如图3所示。从预报结果来看,粒子群-神经网络方法预测基本上能够把异常区域预报出来,预测效果好于逐步回归方法,不足是2022年桂东北预测相反。4 a 预报与实况距平对比符号相同率和均方根误差如表2所示,同号率平均为71.202%,比逐步回归方法高6.645%,比随机森林算法高0.316%;均方根误差平均为0.091,比逐步回归小0.01,比随机森林小0.002。

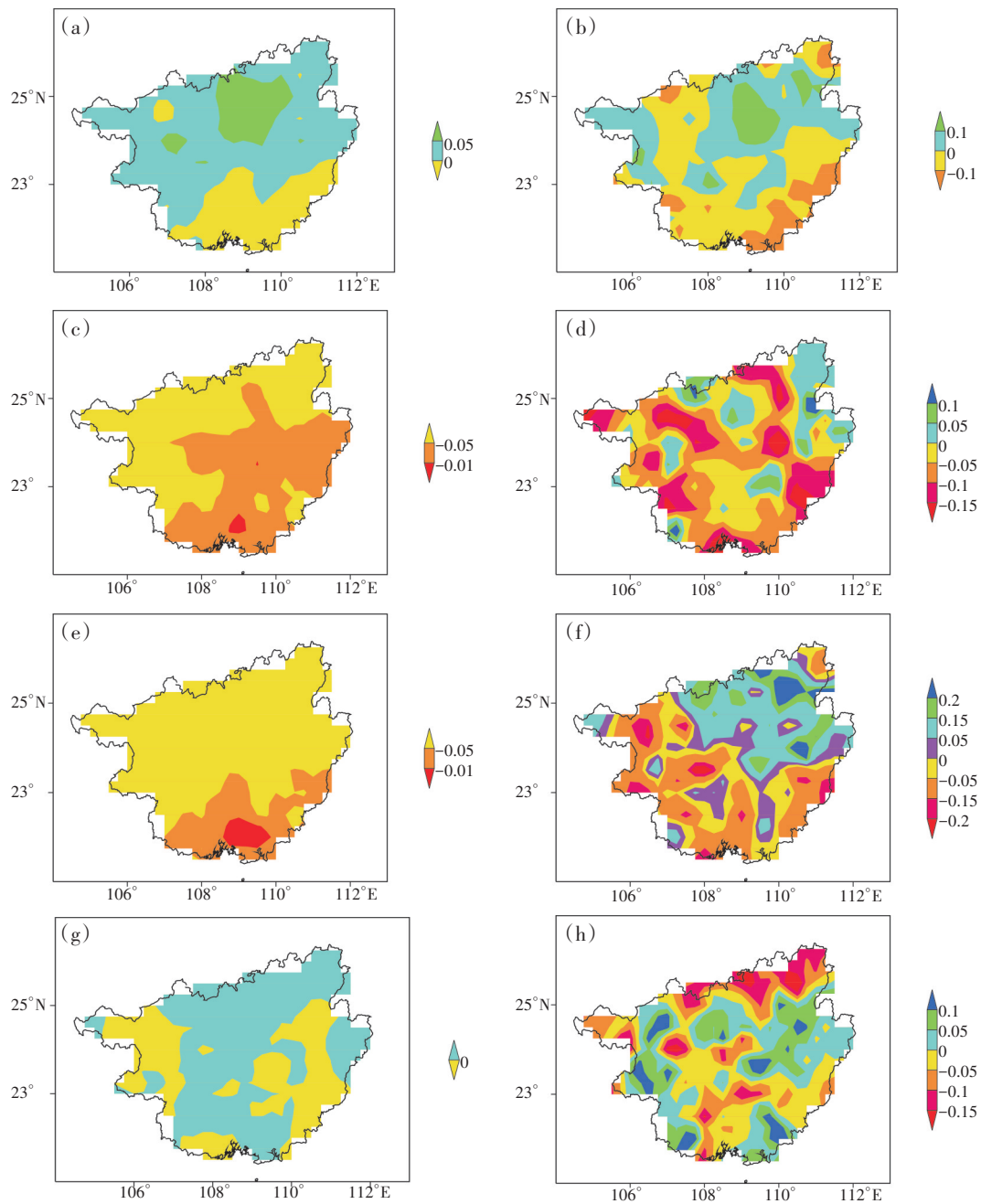


图3 粒子群-神经网络(a、c、e、g)暴雨集中度气候预测图和实况距平图(b、d、f、h)  
(a-b)2020年;(c-d)2021年;(e-f)2022年;(g-h)2023年

### 3.4 基于随机森林算法的暴雨集中度预测和检验

随机森林算法策略树的数量为 50, 构建决策树时对于节点数量和计算量没有限制, 利用最大资源建模直到得到最优解。对 2020—2023 年逐年暴雨集中度距平值进行预测试验, 结果和实况距平对比如图 4 所示。从预报结果来看, 随机森林方法预测基本上能把大部趋势预报出来, 但 2022 年桂东北偏

大、2023 年部分偏大没有预测出来。4 a 预报与实况距平对比符号相同率和均方根误差如表 2 所示, 同号率平均为 70.886%, 比逐步回归方法高 6.329%, 比粒子群-神经网络低 0.316%; 均方根误差平均为 0.093, 大于粒子群-神经网络, 小于逐步回归(表 2)。可见, 利用机器学习算法等人工智能方法建模预测, 能够明显的提高暴雨集中度气候预测能力。

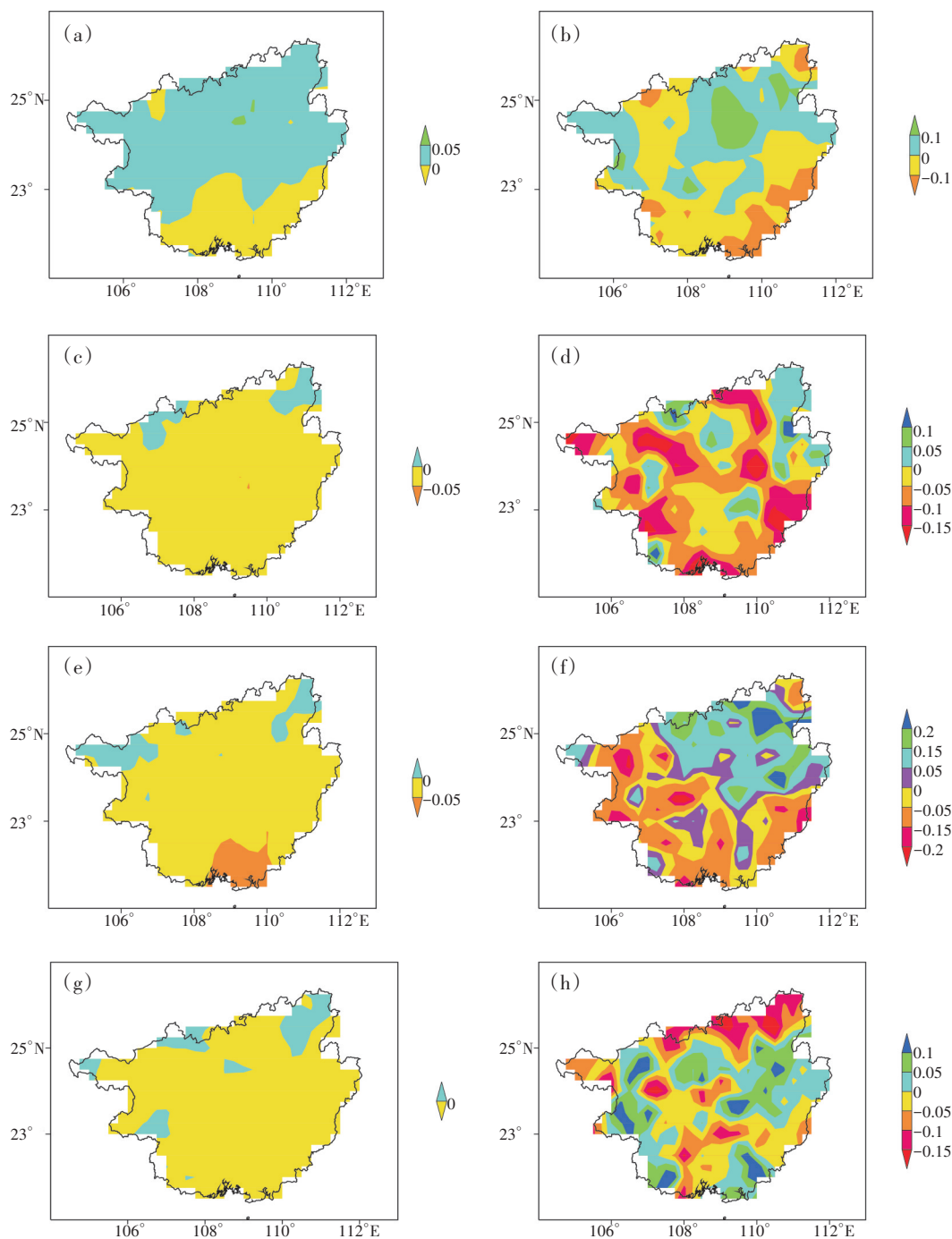


图4 随机森林算法暴雨集中度气候预测图(a、c、e、g)和实况距平图(b、d、f、h)  
(a-b)2020年;(c-d)2021年;(e-f)2022年;(g-h)2023年

## 4 结论和讨论

本研究利用广西气象观测站逐日降水和国家气候中心大气环流、海温指数资料, 构建广西暴雨集中度计算方法, 建立基于逐步回归和机器学习算法的暴雨集中度气候预测模型, 并进行预测试验。得到以下结论:

(1) 广西存在三个暴雨集中度高值区: 以桂林和柳州二市北部为中心的桂东北地区、以“东兰、巴马、凤山”为中心的桂西山区及沿海地区。1994、1998年暴雨集中度最大, 广西均发生严重的暴雨洪涝灾害; 暴雨集中度最小的年份为1963年, 发生严重的干旱灾害。

(2) 建立基于逐步回归方法、粒子群-神经网络、随机森林算法的暴雨集中度气候预测模型, 对2020—2023年暴雨集中度距平值进行气候预测试验, 粒子群-神经网络算法预测效果最好, 其次为随机森林算法, 最差是逐步回归方法。

本研究构建暴雨年内非均匀性分布的指标, 尝试开展智能气候预测, 结果表明智能方法的预测与实况距平同号率达到七成以上, 预测效果良好, 具有很好的参考价值, 解决有关这方面气候预测技术从无到有的问题。使用的因子仅限于国家气候中心下发的大气环流和海温指数, 预测因子范围和数量有些偏少, 一定程度影响预测效果, 以后加入高原积雪、北极海冰等资料深入开展相关研究。

### 参考文献:

- [1] ZHANG L J, QIAN Y F. Annual distribution features of the yearly precipitation in China and their interannual variations [J]. *Acta Meteorologica Sinica*, 2003, 17(2): 146–163.
- [2] 张录军, 钱永甫. 长江流域汛期降水集中程度和洪涝关系研究[J]. *地球物理学报*, 2004, 47(4): 622–630.
- [3] 杨金虎, 李耀辉, 王鹏祥, 等. 中国极端强降水事件年内非均匀性特征分析[J]. *自然资源学报*, 2007, 22(4): 623–633.
- [4] 张运福, 胡春丽, 赵春雨, 等. 东北地区降水年内分配的不均匀性[J]. *自然灾害学报*, 2009, 18(2): 89–94.
- [5] QIN W J, LI D L, LEI X M, et al. Analysis on intra-annual inhomogeneity of rainstorm events in Guangxi [J]. *Journal of Tropical Meteorology*, 2014, 20(2): 135–142.
- [6] KARORI M A, LI J P, JIN F F. The asymmetric influence of the two types of El Niño and La Niña on summer rainfall over southeast China [J]. *Journal of Climate*, 2013(26): 4567–4582.
- [7] ZHU Z, LI T, He J. Out-of-phase relationship between boreal spring and summer decadal rainfall changes in southern China [J]. *Journal of Climate*, 2014(27): 1083–1099.
- [8] Nguyen-Le D, Matsumoto J, Ngo-Duc T. Onset of the rainy seasons in the eastern indochina peninsula [J]. *J. Climate*, 2015(28): 5645–5666.
- [9] 章开美, 李丽平, 陈胜东. 华南前汛期极端降水年代际异常及其影响因子分析[J]. *干旱气象*, 2016, 34(1): 64–74.
- [10] 王彦明, 高建芸, 池艳珍. 太平洋海温场不同时间尺度背景下华南前汛期持续性暴雨的统计特征[J]. *大气科学学报*, 2015, 38(1): 126–131.
- [11] 覃卫坚, 李栋梁, 蔡悦幸. 广西大范围暴雨气候异常对太平洋年代际涛动的响应 [J]. *热带气象学报*, 2022, 38(2): 161–170.
- [12] 蔡学湛. 青藏高原雪盖与东亚季风异常对华南前汛期降水的影响[J]. *应用气象学报*, 2001, 12(3): 358–367.
- [13] WANG C, YANG K, LI Y, et al. Impacts of spatiotemporal anomalies of Tibetan Plateau snow cover on summer precipitation in Eastern China [J]. *Journal of Climate*, 2017(30): 885–903.
- [14] 陆虹, 翟盘茂, 覃卫坚, 等. 低温雨雪过程的粒子群-神经网络预报模型[J]. *应用气象学报*, 2015, 26(5): 513–524.
- [15] 覃卫坚, 陆虹, 黄志, 等. 粒子群-神经网络法在广西寒露风日数预报中的应用[J]. *气象与环境学报*, 2015, 31(6): 158–162.
- [16] 覃卫坚, 李耀先, 陈思蓉, 等. 粒子群-神经网络在华南夏季降水短期气候预测中应用研究[J]. *气象研究与应用*, 2015, 36(2): 1–7.
- [17] 孔庆燕, 史旭明, 金龙. 基于粒子群-支持向量机定量降水集合预报方法[J]. *数学的实践与认识*, 2017, 47(5): 219–225.
- [18] 吴建生, 刘丽萍, 金龙. 粒子群-神经网络集成学习算法气象预报建模研究[J]. *热带气象学报*, 2008, 24(6): 679–686.
- [19] 田心如, 蔡凝昊, 张志薇. 基于气象因子及机器学习回归算法的夏季空调负荷预测[J]. *气象科学*, 2019, 39(4): 548–555.
- [20] KIM H L, KIM B H. Flood hazard rating prediction for urban areas using random forest and LSTM [J]. *Journal of Civil Engineering*, 2020, 24(12): 3884–3896.
- [21] KENNEDY J, EBERHART R C. Particle swarm optimization [C]//IEEE proceedings of ICNN'95–international conference on neural network, 1995: 1942–1948.
- [22] ANON. IEEE international conference on neural networks



- [Z].1995.
- [23] 吴建生,刘丽萍,金龙.粒子群-神经网络集成学习算法气象预报建模研究[J].热带气象学报,2008,24(6):679-686.
- [24] 覃卫坚,李耀先,陈思蓉,等.粒子群-神经网络在华南夏季降水短期气候预测中应用研究[J].气象研究与应用,2015,36(2):1-7.
- [25] 覃卫坚,廖雪萍,陈思蓉.延伸期暴雨过程的神经网络预报技术应用初探[J].气象研究与应用,2018,39(4):1-4.
- [26] RIGET J, VESTERSTROM J S. A diversity-guided particle swarm optimizer-the ARPSO [J]. Technical Report 2002-02, Department of Computer Science, University of Aarhus, 2002:345-350.
- [27] 魏一钊,陈军锋.基于随机森林算法的冻融期土壤蒸发预报模型研究[J].水电能源科学,2021,39(4):20-23.
- [28] 覃卫坚,何莉阳,蔡悦幸.基于两种机器学习方法的广西后汛期降水预测模型[J].气象研究与应用,2022,43(1):8-13.

## Research on intelligent climate prediction methods of rainstorm concentration in Guangxi

QIN Weijian, HE Liyang, CAI Yuexing

(Guangxi Climate Center, Nanning 530022, China)

**Abstract:** In this study, using the daily precipitation of 79 meteorological observation stations in Guangxi and the data of the atmospheric circulation indices and sea surface temperature indices data of the National Climate Center from 1961 to 2023, we constructed a method for calculating the rainstorm concentration in Guangxi, and established a climate prediction model for the concentration of rainstorm based on the stepwise regression method, the particle swarm neural network, and random forest algorithm. The results showed that there are three areas of high rainstorm concentration in Guangxi, namely, the northeastern part of Guangxi centered on the northern part of Guilin and Liuzhou, the mountainous area in the western part of Guangxi centered on “Donglan, Bama, Fengshan” area, as well as the coastal area. The anomaly of rainstorm concentration basically reflects the severity of flooding and drought disasters. The climate prediction model of rainstorm concentration based on the stepwise regression method, particle swarm optimization neural network and random forest algorithm is established. According to the climate prediction experiments in 2020—2023, the most effective prediction was made by the particle swarm-neural network algorithm, followed by the random forest algorithm, and finally by the stepwise regression method.

**Key words:** rainstorm; concentration degree; concentration period; particle swarm neural network; random forest algorithm