

黄颖, 金龙, 陆虹, 等. 2019. 基于智能计算的广西冷湿极端天气定性和定量组合预报方法研究 [J]. 大气科学, 43(6): 1424-1440. HUANG Ying, JIN Long, LU Hong, et al. 2019. A Combined Qualitative and Quantitative Prediction Scheme for Cold-Wet Extreme Weather in Guangxi Based on Intelligent Computing [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 43(6): 1424-1440. doi: 10.3878/j.issn.1006-9895.1908.18248

基于智能计算的广西冷湿极端天气定性和定量组合预报方法研究

黄颖¹ 金龙² 陆虹² 黄翠银² 周秀华²

¹ 广西壮族自治区气象科学研究所, 南宁 530022

² 广西壮族自治区气候中心, 南宁 530022

摘要 论文以逐日气温和降水量数据、NCEP/NCAR再分析资料以及预报场资料为基础, 将表征冬季低温冷害的冷湿指数作为预报量, 先利用随机森林方法进行冬季逐日冷湿极端天气定性判别预报分析, 再进一步以粒子群算法为基础的模糊神经网络集成个体生成技术方法, 建立一种新的非线性智能计算定量集成预报模型 (PSO-FNN), 进行了广西冷湿极端天气定量预报模型的预报建模研究。结果表明, 论文提出的这种以不同的智能计算方法构建的定性、定量综合预报分析方法, 比较符合极端天气小概率事件的预报特点, 其中随机森林算法构建的定性预报模型, 对广西冷湿极端天气事件的预报 TS 评分 (Threat Score) 为 0.77, 空报率为 0.23, 漏报率为 0, ETS 评分 (Equitable Threat Score) 为 0.41, TSS 评分 (True Skill Statistic) 为 0.53。而采用粒子群-模糊神经网络方法构建的极端冷湿指数定量集成预报模型比其他线性和非线性预报模型具有更好的预报精度。其中 PSO-FNN 集成预报模型在预报建模样本和独立预报样本个例相同的情况下, 比回归方法的预报平均绝对误差下降了 25% 以上, 比一般的普通模糊神经网络预报平均绝对误差下降了 14.37%。主要原因是因为 PSO-FNN 集成预报模型通过改进集成个体的预报能力和增强集成个体的种群差异性, 提高了集成预报模型的预报精度。因此, 该智能计算集成预报模型的泛化能力显著提高, 预报结果稳定可靠, 为冷湿极端天气客观预报提供了新的预报工具和预报建模方法。

关键词 广西冷湿极端天气 定性、定量组合预报 粒子群-模糊神经网络集成模型 随机森林

文章编号 1006-9895(2019)06-1424-17

中图分类号 P456.1

文献标识码 A

doi:10.3878/j.issn.1006-9895.1908.18248

A Combined Qualitative and Quantitative Prediction Scheme for Cold-Wet Extreme Weather in Guangxi Based on Intelligent Computing

HUANG Ying¹, JIN Long², LU Hong², HUANG Cuiyin², and ZHOU Xiuhua²

¹ Guangxi Institute of Meteorological Sciences, Nanning 530022

² Guangxi Climate Center, Nanning 530022

Abstract Taking the cold-wet index measuring chilling damage in winter as the predictand, a qualitative, discriminant prediction model has been developed for daily cold-wet extreme weather forecasting in winter. This model is based on a

收稿日期 2018-11-02; 网络预出版日期 2019-08-30

作者简介 黄颖, 女, 1983 年出生, 硕士, 高级工程师, 主要从事天气预报技术方法研究与业务应用工作。E-mail: yinger2001@126.com

通讯作者 金龙, E-mail: jinlong01@163.com

资助项目 国家自然科学基金项目 41565005、41575051, 广西自然科学基金项目 2018GXNSFAA281229、2017GXNSFDA198030

Funded by National Natural Science Foundation of China (NSFC) (Grants 41565005, 41575051), Guangxi Natural Science Foundation (Grants 2018GXNSFAA281229, 2017GXNSFDA198030)

Random Forest (RF) algorithm, and uses daily temperature and precipitation data, NCEP/NCAR reanalysis data, and 24-h forecast field data. Further, a new, nonlinear, intelligent computing, quantitative ensemble prediction scheme has been developed for predicting cold-wet extreme weather in Guangxi by employing Particle Swarm Optimization (PSO) algorithm, and it is termed the PSO-FNN ensemble prediction model. The ensemble members of the PSO-FNN model were generated by adopting the PSO algorithm. Results showed that the qualitative and quantitative comprehensive prediction based on different intelligent computing methods proposed in this paper were in accord with the forecast characteristics of the small probability extreme weather event. Threat score of the qualitative forecast model based, on the RF algorithm for cold-wet extreme weather in Guangxi, was 0.77, false alarm rate was 0.23, missing rate was 0, equitable threat score was 0.41, and the true skill statistic was 0.53. Moreover, the forecast accuracy of the quantitative PSO-FNN ensemble prediction model was higher than those of the linear and nonlinear forecast models. Using identical modeling samples and independent samples, the PSO-FNN ensemble prediction model showed the reduction in mean absolute errors being >25% relative to the stepwise method, and 14.37% relative to the normal fuzzy neural network. Analyses of the new scheme suggested that the forecast accuracy of the ensemble prediction model was improved by enhancing the prediction ability and population diversity of the individual ensemble members. Therefore, the generalization capacity of the intelligent computing ensemble prediction model was significantly enhanced, and the forecast results were stable and reliable, providing new forecasting tools and prediction modeling methods for objective forecasts of cold-wet extreme weather.

Keywords Cold-wet extreme weather in Guangxi, Qualitative and quantitative combined prediction, PSO-FNN ensemble model, Random forest

1 引言

近年来, 北半球区域性极端低温事件频繁发生, 如2008年1月中国南方地区发生了历史上罕见的大范围低温雨雪冰冻灾害; 2009~2011年冬季, 包括中国在内的北半球中高纬度地区经历了持续发生大范围低温冰雪异常事件, 尤其是2011年1月中国贵州出现仅次于2008年初的低温雨雪冰冻天气, 过程长达32天(杜小玲等, 2014)。尤为引起关注的是从2012年1月下旬开始的持续性极寒天气, 席卷了整个欧亚大陆, 欧美媒体甚至称此事件为“小冰期”来临的征兆(布和朝鲁等, 2018)。由于极端天气事件具有持续时间长、空间范围广和强度大等特点, 给经济发展、人民生活造成了重大影响。因此, 迫切需要针对持续性区域极端低温冷害天气, 提出更为有效的客观预报新方法, 以提高区域极端气象灾害的防御能力。

目前, 冷湿极端天气事件引起国际和国内气象学者的广泛关注, 已经取得了一系列研究成果(Barszcz et al., 2018; Ressler et al., 2012; Tobin and Kumjian, 2017; Qian et al., 2014)。丁一汇等(2008)、李崇银等(2008)、陶诗言和卫捷(2008)、彭艳等(2010)、赵思雄和孙建华(2008)、王东海等(2008)及 Bao et al. (2010)等

学者从不同角度对2008年初中国南方的低温雨雪冰冻灾害天气的形成机理及其致灾原因进行分析, 发现其主要成因是由于欧亚大陆出现异常的大气环流, 多种因素相互配合和迭加, 致使出现了严重的持续性冷湿天气。在区域性极端事件识别方面, Ren et al. (2012)在客观天气图分析法的基础上, 提出了持续性区域极端事件客观识别方法。龚志强等(2012)对该方法进行完善, 发展了一种针对区域性极端低温事件的客观识别方法。王晓娟等(2013)根据事件的空间分布特征, 提出了基于区域性极端低温事件客观识别技术。而在冰冻雨雪的预报研究方面, 钱维宏和张宗婕(2012)通过对2008年初的南方低温雨雪冰冻天气事件的研究, 从天气扰动物理分解方法中获得前期信号, 预报了2011年初发生在中国南方地区的冻雨事件。江漫等(2014)对全球大气变量再分析资料做物理分解得到天气尺度扰动分量用于低温雨雪冰冻事件的早期信号分析, 最多可提前10天追踪到。Qian et al. (2014)制定了单站和区域出现持续性低温雨雪事件的标准。陆虹等(2015)以低温雨雪过程的冷湿程度及影响范围为标准, 将低温雨雪过程分为一般过程和严重过程, 进行预报建模研究。这些方法在预报方法研究和预报试验方面都取得了较好的成果。

由于极端天气事件由多个天气系统在同一时段、同一地区相互作用及影响造成,其变化有很强的无序性、非线性和时变性特征,而神经网络、遗传算法、粒子群算法以及模糊神经网络等非线性智能计算方法因具有较强的处理非线性问题的能力,近年来在包括大气科学在内的很多学科中得到越来越多的应用(黄颖等, 2014; 金龙, 2004; Huang et al., 2018; Jin et al., 2015; Jin et al., 2008)。金龙等(2003)采用模块化模糊神经网络方法进行了一种新的数值预报产品释用预报研究。周岩等(2010),何伟等(2004)将粒子群优化算法与模糊神经网络融合进行降水、气温的短时预报和径流预报。衣治安等(2014)提出了一种基于混合改进粒子群算法的模糊神经网络的管道腐蚀动态检测方法。师彪等(2010)采用改进的粒子群—模糊神经网络混合优化算法进行短期电力负荷的预报研究。

由于一般的单个模糊神经网络模型其初始权值是由随机产生(每次会有差异),即使采用粒子群等优化算法对权值和隶属度函数作优化计算,但其本质仍是单个的模糊神经网络模型。因此,为了提高预报方法稳定性和确定性,本文尝试进行一种首先运用随机森林方法对冷湿极端天气事件进行定性判别预报(仅作未来24 h是否出现极端冷湿天气的定性判别预报),再进一步仿照数值预报的集合预报思想,利用进化计算的粒子群算法生成期望输出相同的 N 个模糊神经网络个体,结合主成分分析的预报因子信息特征提取技术,构建一种由几十个模糊神经网络预报模型预报得出的一种新的非线性粒子群—模糊神经网络冷湿事件定量集成预报模型,通过这种基于不同智能计算预报建模方法对广西的24 h冷湿极端天气进行基于智能计算方法的定性、定量组合预报研究。

2 预报量和预报因子的计算分析

2.1 预报对象(冷湿指数)

冬季的冷害影响除了气温有最重要作用外,与此同时如果叠加有降水的湿度增加,则冷害程度会越来越显著,为此,关于冷湿指数的研究已引起很多气象学者的关注(龚志强等, 2012; 万素琴等, 2008; Qian et al., 2014; Zhang et al., 2011)。Zhang et al. (2011)利用逐日气温和降水数据定义了单日冷湿指数(PT),计算公式如下:

$$PT_i = \frac{P_i - \bar{P}}{P_s} - \frac{T_i - \bar{T}}{T_s}, \quad (1)$$

其中, P_i 、 T_i 分别为过程第 i 日的降水量和平均温度, \bar{P} 、 \bar{T} 分别为历史同期降水量的平均值和平均温度的平均值, P_s 、 T_s 分别为历史同期降水量的标准差和平均温度的标准差。 PT 值越大,单日冷湿特征越明显。

为了更准确地反映我国华南地区的冷害程度,周秀华等(2019)通过分析发现,由于原冷湿指数计算公式(1)没有考虑我国华南地区冬季日降水量远比北方地区大,而增大的降水量会明显夸大实际的低温冷害程度,因而改进并提出了一种将降水量的影响作分级处理的方法,更适用于我国华南地区的冷湿指数的计算,本文采用这种新方法计算冷湿指数作为预报量进行预报建模研究,具体公式如下(周秀华等, 2019):

$$\begin{cases} PT_i = \frac{\bar{P} + a}{P_s} - \frac{T_i - \bar{T}}{T_s}, & P_i \geq \bar{P} \\ PT_i = \frac{P_i}{P_s} - \frac{T_i - \bar{T}}{T_s}, & 0 < P_i < \bar{P} \\ PT_i = \frac{-\bar{P}}{P_s} - \frac{T_i - \bar{T}}{T_s}, & P_i = 0 \end{cases} \quad (2)$$

在公式(2)中,若 $T_i \leq 2^\circ\text{C}$, $a=1$;若 $2 < T_i \leq 4^\circ\text{C}$, $a=0$ 。在计算冷湿指数时,本文选取我国广西($21^\circ \sim 26.50^\circ\text{N}$, $104.35^\circ \sim 112.20^\circ\text{E}$)为研究区域,以1956~2017年每年12月至次年2月广西92个国家级地面气象站的逐日平均气温和降水量数据为基础(研究区域和站点分布如图1所示),根据公式(2)以达到冷湿指数计算标准站点的PT指数累加值作为本研究的预报量。

通过对1956~2017年每年12月至次年2月(即1956~2018年)广西区域内低温雨雪天气进行的计算分析,得到区域持续性冷湿过程42次,累计天数共332 d。本文以研究区域内这332 d达到持续性冷湿事件标准的站点冷湿指数之和($\sum PT$)作为预报量序列,进行基于智能计算的冷湿极端天气定性、定量组合预报方法建模研究。

2.2 低温冷湿预报建模的预报因子计算

自2008年初中国南方发生罕见的低温雨雪冰冻天气以来,气象学者愈加关注对这类灾害性天气的研究,从气候背景、大气环流异常、动力学条件、锋区特征等方面进行了研究,得到了一些有意

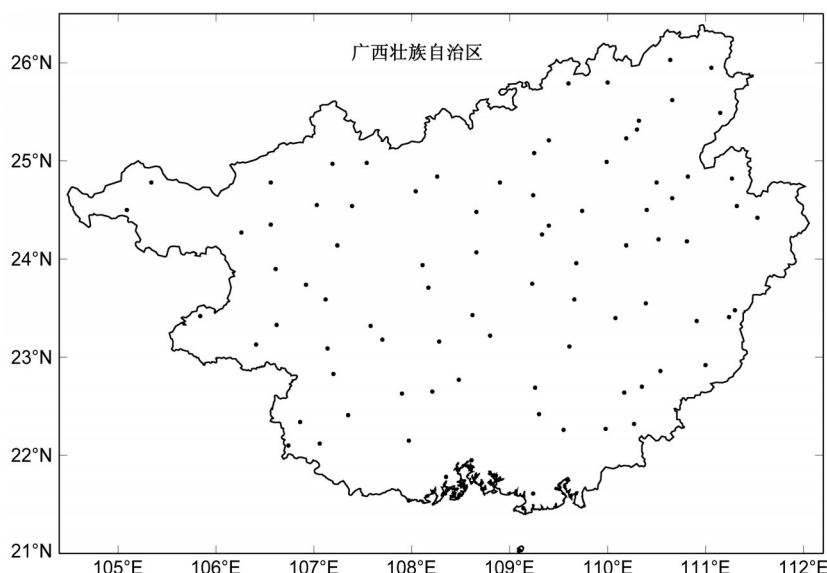


图1 冷湿极端天气研究区域和广西气象站分布
Fig.1 Research area for cold-wet extreme weather and distribution of meteorological stations in Guangxi

义的成果。例如，对于2008年初雨雪冰冻事件，李崇银等（2008）指出区域持续性低温雨雪事件的主要原因是阻塞高压、副热带高压、南支槽等多个大气环流系统的异常配合造成的。丁一汇等（2008）分析发现欧亚大气环流异常持续性是造成冷空气不断侵袭中国南方的直接原因，孟加拉湾和南海地区暖湿气流的北上是大范围冻雨和降雪形成并持续在中国南方的必要条件。陶诗言和卫捷（2008）指出欧亚大陆大气环流异常，亚洲中高纬度阻塞高压稳定维持达20余天，为中国南方暴雪冻雨过程提供了环流背景场。王东海等（2008）研究发现阻塞上游50°N区域有极强的负涡度平流持续输送到阻塞区，使濒临崩溃的阻塞形势得以重新加强，从而使阻塞形势长时间稳定维持。杜小玲等（2014）对2011年1月发生在贵州的低温雨雪冰冻天气进行了分析，分析表明维持在25°N附近的低层切变为低温雨雪天气的发生提供了辐合条件，地面上稳定持久的准静止锋是低温雨雪天气发生的重要影响系统。

根据以往的研究结果，以挑选造成广西区域持续性冷湿事件相关性高、物理意义明确的物理量场为前提，将NCAR/NCEP资料925、850、700、500、200 hPa高度场、温度场、湿度场、涡度场、散度场、垂直速度场、水平和垂直风场，及海平面气压场等再分析资料（Kalnay et al. 1996），共35个物理量场作为基本预报因子场，普查计算分析

332 d区域持续性冷湿天气事件与前期各预报因子场的相关关系。2011~2018年39个独立预报样本对应的因子为24 h预报场的环流资料。其中，NCAR/NCEP再分析资料的分辨率为 $2.5^{\circ} \times 2.5^{\circ}$ ，经度范围： $0^{\circ} \sim 180^{\circ}$ ，纬度范围： $20^{\circ}S \sim 80^{\circ}N$ ，共 41×73 个格点。具体做法如下：以冷湿指数（PT值）与前期预报因子场成片相连的相关格点，且相关系数绝对值 $|R| \geq 0.15$ （达到0.01显著性水平）作为1个预报因子区，并在该预报因子区内挑选出相邻成片的相关系数绝对值 $|R| \geq 0.20$ （达到0.001显著性水平）的相关格点作为一个预报因子的计算格点区。这样的计算规定，可以保证每一个前期预报因子是由1个大范围的相关区域得出，以增强预报因子的稳定性，而从中再进一步选出更高相关的格点区域，则可以保证预报因子的高相关性。由于将两个相邻或相近区域且符号相反的因子进行相加或相减后得到的组合因子，其相关系数比单一因子高。例如，在地面温度场挑选出5个因子的相关系数分别为-0.4620、0.2272、0.2616、0.2646、0.2258（见图2），将第1个因子分别与第2至第5个因子进行相减组合，得到组合因子的相关系数分别为-0.5073、-0.5207、-0.5431、-0.5181，相关系数均有较大的提高。因此，在挑选因子时，采用这种组合因子，可以使模型输入包括更多的有效预报信息，增强预报模型的物理基础。

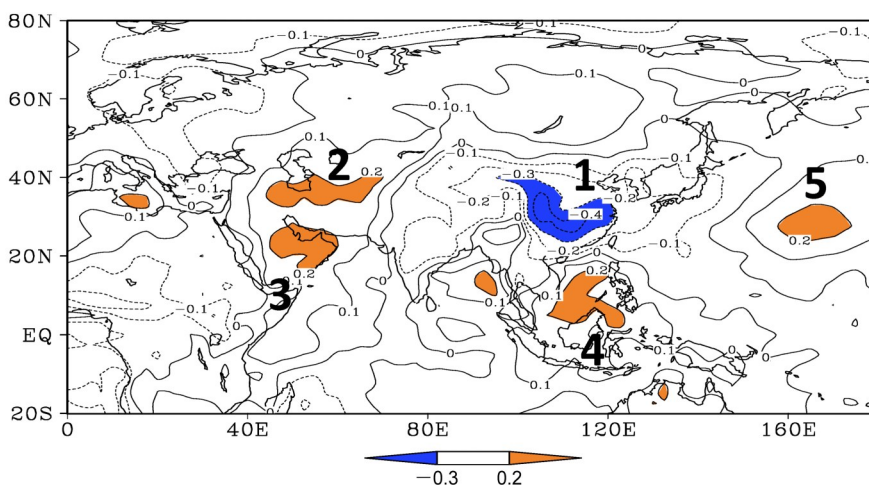


图2 冷湿指数 (PT) 与地面温度场相关分布 (阴影表示达到0.001显著性水平, 数字1至5分别表示5个高相关区)

Fig. 2 Correlation coefficients between PT (cold-wet index) values and ground temperature field (shaded area reached the level of 0.001 significance, and numbers 1-5 denote five high correlation regions)

经相关普查计算分析, 以1956~2017年每年12月至次年2月332 d持续性冷湿事件为预报量, 共提取出270个相关系数较高、具有冷湿事件预测强信号天气特征的因子作为初选因子, 其中, 高度场因子21个, 温度场因子27个, 湿度场因子28个, 风场(经向风和纬向风)因子87个, 涡度场因子44个, 散度场因子38个, 垂直速度场因子20个, 海平面气压场5个; 140个为组合因子, 130个为单因子; 因子相关系数的绝对值($|R|$)范围在 $[0.20, 0.60]$ 之间(具体因子表略)。本文以这270个因子作为模型输入, 对影响广西的332d持续性冷湿事件进行随机森林算法的冷湿极端天气定性判别预报研究。

3 冷湿指数的定性判别预报方法

由于我国华南区域大范围、长时间的低温冷湿极端天气过程是小概率事件, 是否可以先对未来24 h低温冷害天气过程出现的可能性作定性的判别预报后, 再作进一步的定量客观预报, 来提高对小概率事件的极端天气过程的预报能力。为此, 本文尝试先采用包含多个决策树的随机森林分类算法, 开展随机森林算法的冷湿事件定性判别预报建模研究。

3.1 随机森林分类算法

随机森林(Random Forest, 简称RF)算法由Breiman(2001)提出, 是一种非线性统计方法, 利用自助法(Bootstrap)从原始样本中有放回地随

机抽取多个训练样本子集, 对每个样本子集分别进行决策树建模。对于分类问题, 按多棵决策树分类器投票决定最终分类结果, 其本质与气象上的集合预报思想相近, 综合考虑多棵决策树模型的预测结果进行集成, 并且每一棵决策树之间是没有关联的, 避免了过拟合问题。该方法具有学习过程快速、稳定性好、预测精度高的优点, 并已在生物、医学等多个学科领域得到了较成功的应用研究(李建更等, 2009; 李欣海, 2013; 邹亮等, 2012)。

近年来, 随机森林算法也引起了气象学者的关注。例如, Herman and Schumacher(2018)利用随机森林方法进行极端降水的预测。沈润平等(2017)构建了基于随机森林遥感干旱监测模型。余胜男等(2016)采用随机森林方法对长江中下游地区1月降水量进行长期预报。张晓羽等(2016)通过建立随机森林模型分类法对森林植被类型进行分类。赵华生等(2018)提出一种基于最大相关最小冗余度和随机森林回归算法的ECMWF集合预报产品暴雨预报释用方法。刘吉敏等(2018)采用随机森林方法对西北太平洋变性热带气旋变性完成时间进行判定。而本文则尝试先采用随机森林方法对未来24 h是否出现极端冷湿天气的可能性, 作出有无可能的定性判断预报, 然后再进一步利用粒子群-模糊神经网络集成预报方法对24 h极端冷湿天气进行定量预报。

随机森林方法在构建训练样本子集上采用自助法(Bootstrap)抽样。具体做法如下: 对样本量为

N 的样本集 S 做 N 次有放回的随机重复抽样 m 次，得到 m 个训练样本集 S_i ($i=1, 2, \dots, m$)，并在此基础上建立 m 棵决策树预测模型 T_i ($i=1, 2, \dots, m$)。其中，在利用训练样本生成决策树时，每个节点都将随机选择 k ($k < K$, K 为输入变量数) 个特定的变量来确定最佳的分裂点，并且每棵决策树都最大可能地进行生长而不进行剪枝。最后，根据 m 棵决策树得到的分类结果 $Y=\{Y_1, Y_2, \dots, Y_m\}$ ，通过投票法，得到最多票数的类别作为最终的模型输出。

3.2 冷湿极端天气的定性判别预报试验

由于随机森林算法可以高效地评估所有变量的重要性，不需要考虑一般回归分析面临的多元共线性的问题 (李欣海, 2013)。因此，本文直接以基于 NCAR/NCEP 资料相关普查得到的 270 个预报因子中相关系数的绝对值 $|R| > 0.4$ 的因子 (达到 0.001 显著性水平) 作为模型输入，利用随机森林分类方法可以随机选择决策树节点划分特征，高度并行训练数据的能力，构造了一种基于随机森林方法的冷湿极端天气定性判别预报模型。

本文将 $PT \geq 30$ 作为冷湿指数极端天气过程分类阈值，以影响广西的 332 d 冷湿事件为预报对象，将 PT 值分成两类，一类是强冷湿事件 ($PT \geq 30$)，记为“1”；另一类是弱冷湿事件 ($PT < 30$)，记为“0”，这样就把 PT 值排列为 01 序列。根据李欣海 (2013) 研究表明，当随机森林算法的决策树数量达到 > 60 时，随机森林的预测误差较小，并趋于稳定。因此，本文随机森林模型的决策树数量取 75，以 1956~2010 年 293 个样本为建模样本，2011~2018 年 39 个样本为独立预报样本，进行冷湿极端天气事件的定性判别预报试验。以预报第 1 个独立样本为例，随机森林定性判别计算过程的第 1 和第 2 棵决策树模型见图 3，其中“1”判定为 24 h 后出现强冷湿极端天气，“0”为弱冷湿极端天气。

采用随机森林方法进行冷湿极端天气事件的定性判别预报试验时，以预报第 1 个独立样本为例，首先从 270 个自变量中挑选出与预报量相关系数 $|R| > 0.4$ 的 65 个因子，以这 65 个因子作为随机森林定性判别模型输入，采用 Bootstrap 方法从 293 个建模样本中随机的有放回的抽样生成 75 组训练样本集，并在此基础上建立 75 棵决策树预测模型，其中第 1 棵决策树模型见图 3a。该模型从 65 个自变量中选择部分变量作为决策树节点划分特征。第

1 个节点为第 35 号因子，第 2 个节点分别为第 44 和第 10 号因子，经过不断分裂构造决策树的各个分支，直到这棵树能够准确表示分类或者遍历所有属性，构建出第 1 棵决策树模型。利用第 1 个独立样本的因子作为这棵决策树模型输入，得出冷湿天气判定结果 Y_{11} 。第 2 棵决策树的第 1 个节点为第 4 号因子，第 2 个节点分别为第 33 和第 9 号因子，不断分裂后建立了第 2 棵决策树 (图 3b)，输入因子得到判定结果 Y_{12} 。依此类推，得到冷湿指数第 1 个独立预报样本的 75 棵决策树的判定结果 $Y_1=\{Y_{11}, Y_{12}, \dots, Y_{1,75}\}$ 。通过投票法，取得冷湿指数第 1 个独立预报样本的判定预报结果 Y_1 。在预报冷湿指数第 2 个独立样本时，将 1~293 个样本与第 1 个独立样本作为建模样本，按照同样的建模方法，取得第 2 个独立预报样本的判定预报结果 Y_2 。依此类推，最终获得 39 个冷湿指数独立预报样本的判定结果 $Y=\{Y_1, Y_2, \dots, Y_{39}\}$ ，其中 39 个独立预报样本对应的随机森林模型输入因子数为 46~65 ($|R| > 0.4$)。由图 3 可以看出，由于 Bootstrap 抽样产生的训练样本集和划分决策树节点的自变量及划分阈值是不相同的，因此构建的 75 棵决策树模型都可能是不一样的，并且随机森林算法在决策树生长过程中，不对其进行修剪，保证每棵决策树最大限度的生长，以使决策树具有高度的差异性，保证投票结果的客观性和稳定性。

由于目前尚未有专门的冷湿事件定性判别预报的检验标准，为了客观评价基于随机森林算法的冷湿极端天气定性判别分析方法的优劣性，本文采用与降水 TS 评分 (Threat Score)、ETS 评分 (Equitable Threat Score) 及 TSS 评分 (True Skill Statistic) 等相同的评分检验方式，对定性判别预报结果进行评定，各评分计算公式如下：

$$TS = \frac{H}{H + F + M}, \quad (3)$$

$$FAR = \frac{F}{H + F}, \quad (4)$$

$$PO = \frac{M}{H + M}, \quad (5)$$

$$ETS = \frac{H - \hat{R}}{H + F + M - \hat{R}}, \quad (6)$$

$$\hat{R} = \frac{(H + F)(H + M)}{H + F + M + Z}, \quad (7)$$

$$TSS = \frac{H}{H + M} - \frac{F}{F + Z}, \quad (8)$$

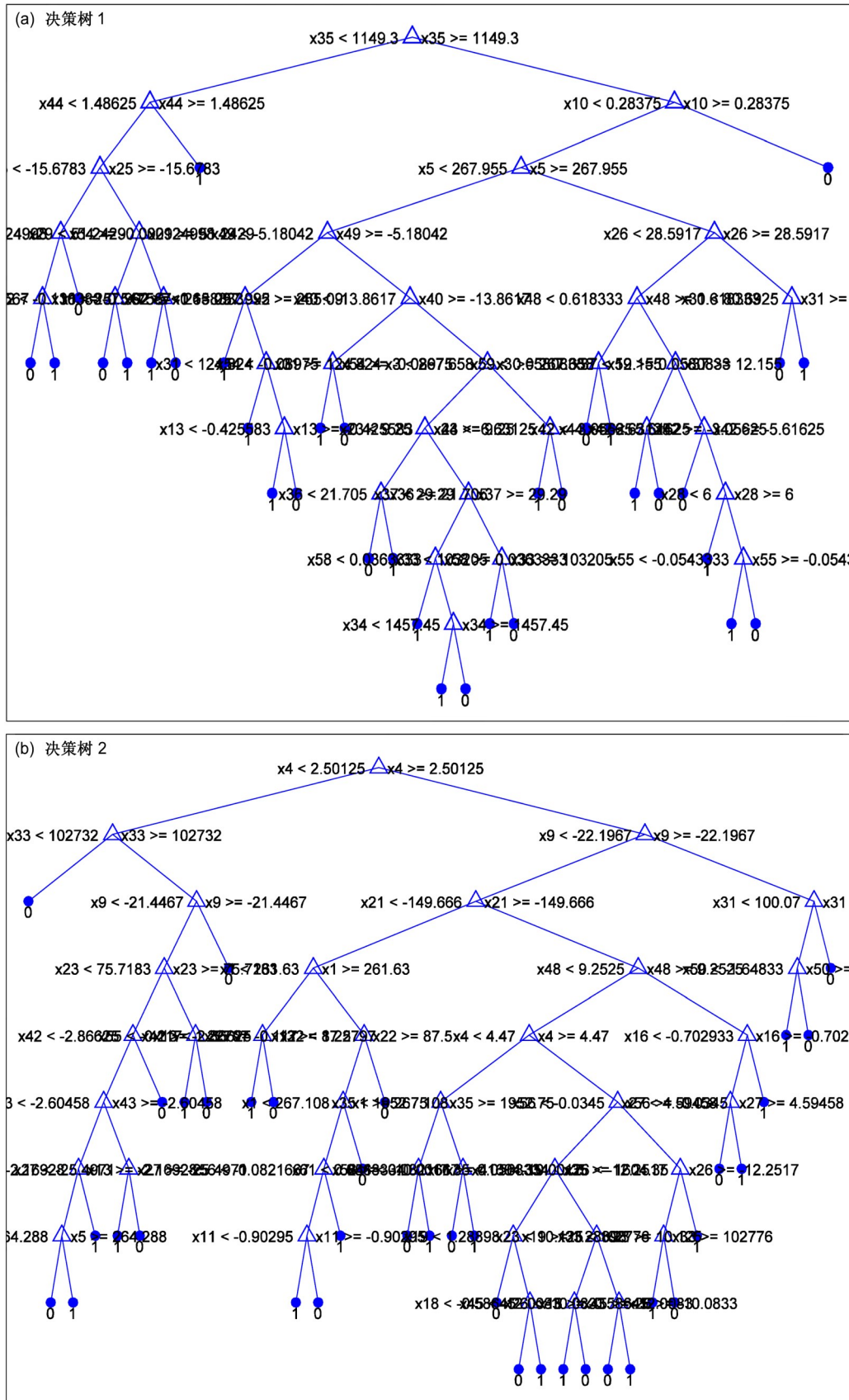


图3 随机森林定性判别计算第1个独立预报样本的 (a) 第1棵和 (b) 第2棵决策树模型

Fig.3 The (a) 1st and (b) 2nd decision trees for the 1st independent prediction sample using qualitative discrimination calculation of a random forest

其中， H 、 F 、 M 分别是正确、空报、漏报的次数； FAR 、 PO 分别表示空报率、漏报率； Z 为报对弱冷湿极端天气事件的次数。

公式(6)中，ETS评分可以有效去除随机预报对评分的影响，取值范围是 $-1/3 \sim 1$ 。当无空报、漏报时， $ETS=1$ ，此时预报效果最佳；当 $ETS \leq 0$ 时，表示无预报技巧。

公式(7)中，TSS评分用于全面评价某级预报的总技巧，不但包括强冷湿事件的预报技巧，还包括对一般强度冷湿事件的预报技巧。当无空报、漏报时， $TSS=1$ ，预报效果最好；当强冷湿事件和一般强度冷湿事件的预报都不准确时， $TSS=-1$ ，此时没有预报技巧；而如果小概率事件预报准确，TSS评分也会较高。

经统计，在独立预报试验中， $H=24$ ， $F=7$ ， $M=0$ ， $Z=8$ ，因此，TS评分为0.77，空报率为0.23，漏报率为0，ETS评分为0.41，TSS评分为0.53。各评分计算结果表明，判定结果稳定，具有较高的预报技巧，并且随机森林方法判别出全部 $PT \geq 30$ 冷湿事件(24次)，特别地，该方法也准确地对4次 $PT \geq 100$ 的强冷湿天气作出了准确的判别预报。这是因为随机森林模型通过对大量决策树的汇总提高了模型的预测精度，因此在冷湿事件定性判别预报具有较好的预报效果。

由于这种基于随机森林方法的定性预报模型，只是对未来是否出现强冷湿天气事件作出了较为准确的趋势预报，然而实际的灾害性天气预报是需要对这种灾害性天气的强度作出客观的定量预报。为此，本文采用基于粒子群—模糊神经网络集成预报方法，进一步进行冷湿过程强弱程度的客观定量预报。尝试采用先定性后定量的综合预报分析方法，以提高这种发生概率较低的强冷湿天气事件的预报准确率。

4 粒子群—模糊神经网络集成的广西强冷湿指数的定量预报方法

本文在设计考虑进行冷湿指数的客观定量预报建模时，为了增强预报模型的稳定性和准确性，仿照数值天气预报模式的集合数值预报方法，采用粒子群算法，同时构建期望输出相同的 N 个模糊神经网络模型，最终的定量预报结果则采用这 N 个模糊神经网络的预报结果作集成，进行24 h持续性强冷

湿指数天气的定量集成预报。

4.1 粒子群—模糊神经网络集成预报模型原理

4.1.1 基本模型

在设计利用粒子群算法构建粒子群—模糊神经网络集成预报模型的预报个体时，模糊神经网络的基本模型是采用数值型四层前馈网络(王士同, 1998)，依次为输入层、隶属度生成层、模糊推理层和反模糊化输出层，包含了模糊系统中模糊化—模糊推理—解模糊的基本模块，每层节点数及权值可根据模糊系统所采用的具体模块形式而预置。模糊神经网络结构的各层说明如下(王士同, 1998; Huang et al., 2018)：

(1) 第1层：输入层。每个节点 x_i ($i=1, 2, \dots, n$, n 为输入变量数)表示一个输入变量。

(2) 第2层：隶属度生成层，即为模糊化层。采用高斯函数，计算输入层的输入变量隶属于模糊集的隶属度，表达式为

$$\mu_{ij}(x_i) = \exp\left[-\frac{(x_i - a_{ij})^2}{\sigma_{ij}^2}\right],$$
$$i = 1, 2, \dots, n \quad j = 1, 2, \dots, m, \quad (9)$$

其中， μ_{ij} 是 x_i 的第 j 个隶属度函数， a_{ij} 和 σ_{ij} 分别是隶属度函数的中心和宽度， m 为模糊规则数。

(3) 第3层：模糊推理层。对模糊逻辑“and”运算按照相乘方式，将各节点的输出值定义为该节点所有输入的代数乘积：

$$\pi_j = \mu_{1j} \times \mu_{2j} \times \dots \times \mu_{nj} = \prod_{i=1}^n \mu_{ij}, \quad j=1, 2, \dots, m \quad (10)$$

(4) 第4层：反模糊化输出层。表示为所有输入信号的叠加：

$$\hat{y} = \omega_1 \pi_1 + \omega_2 \pi_2 + \dots + \omega_m \pi_m, \quad (11)$$

其中， ω_j ($j=1, 2, \dots, m$)为第 j 个模糊神经元的权值。

4.1.2 粒子群—模糊神经网络集成算法

粒子群(Particle Swarm Optimization, 简称PSO)优化算法是通过模拟鸟群觅食行为而发展起来的一种基于群体协作的随机搜索进化算法，具有较强的全局搜索能力(Kennedy and Eberhart, 1995)。粒子群算法首先初始化一群随机粒子，每个粒子都代表着优化问题在整个搜索空间的一个可能解，每个粒子的优劣由目标函数计算的适应度值决定，并且根据粒子搜索到的个体极值和全局极值进行更新，通过追随当前的最优粒子迭代搜索找到

最优解。本文利用粒子群算法的全局搜索能力确定模糊神经网络结构，其中适应度计算公式定义为

$$F_{\text{PSO}} = \frac{1}{\sqrt{\frac{1}{N} \sum_{j=1}^n (y_j - \hat{y}_j)^2}}, \quad (12)$$

其中， \hat{y}_j 为模糊神经网络的实际输出， y_j 为模糊神经网络的期望输出， N 为训练样本数。该适应度函数保证网络的实际输出与期望输出之间的误差越小，适应度值越大。

在粒子群算法中，粒子速度的进化方程为 (Shi and Eberhart, 1998):

$$v(t+1) = w \cdot v(t) + c_1 r_1 [\text{pbest}(t) - x(t)] + c_2 r_2 [\text{gbest}(t) - x(t)], \quad (13)$$

其中， v 是粒子的飞行速度； x 是粒子的位置； t 是时刻； c_1 、 c_2 为学习因子； r_1 、 r_2 为(0, 1)间的随机数。为了平衡粒子群算法的全局和局部搜索能力，惯性权重 w 采用下式进行自适应调整：

$$w(t) = w_{\text{max}} - \frac{w_{\text{max}} - w_{\text{min}}}{\text{iter}_{\text{max}}} \cdot \text{iter}, \quad (14)$$

其中， iter 、 iter_{max} 分别是当前迭代次数和最大迭代次数； w_{max} 、 w_{min} 分别是惯性权重的最大值和最小值 (黄颖等, 2014; Kennedy and Eberhart, 1995)。

粒子群—模糊神经网络集成算法的具体计算步骤如下：

(1) 初始化粒子群。由第 4.1.1 节可以看出，与传统误差反向传播 (BP) 神经网络不同，模糊神经网络训练调整的是模糊规则数、模糊化中心、宽度及反模糊化输出层的连接权值。因此，随机生成 N 个粒子个体的位置矩阵和速度矩阵，代表了 N 个模糊神经网络预报个体。其中，位置矩阵对应模糊神经网络结构，控制了网络参数的大小。

(2) 计算适应度。对当前粒子群解码得到 N 个模糊神经网络，输入训练样本，根据公式 (9) 至 (11) 得到各模糊神经网络个体的输出。将预报输出代入式 (12)，计算出粒子的适应度。

(3) 初始化个体极值 pbest 及全局极值 gbest 。

(4) 评价每个粒子。根据粒子适应度更新个体极值 pbest 和全局极值 gbest 。

(5) 根据公式 (13) 和 (14) 更新粒子速度。

(6) 粒子位置的更新。位置矩阵中，粒子的位置 x 的进化方程为

$$x(t+1) = x(t) + v(t+1). \quad (15)$$

(7) 反复进行公式 (2) 至 (6) 的迭代计算，直到满足最小训练误差或当前的迭代次数达到了最大迭代次数，则训练结束。

将最后一代的 N 个粒子全部解码，以其作为 N 个模糊神经网络的初始网络结构，再次输入训练样本，由此得到 N 个模糊神经网络预报模型。本文采用算术平均法对这 N 个模糊神经网络预报成员进行集成预报建模，作为粒子群—模糊神经网络集成预报模型的预报值。图 4 给出了广西冷湿极端天气的随机森林算法的定性预报与粒子群—模糊神经网络定量组合预报的预报计算流程图。

4.2 广西强冷湿指数的极端天气定量预报试验

由于严重的雨雪冰冻灾害会给人民的生产和生活造成巨大损失，而这种严重的强灾害性天气发生概率低，是天气预报的重点和难点。为此，本文从第 2 节的 1956~2018 年影响广西的 332 d 持续性冷湿事件中，以 $\text{PT} \geq 50$ 的强冷湿事件共 131d 作为预报量，将 1956~2010 年 $\text{PT} \geq 50$ 的强冷湿事件共 115d 作为预报建模样本，以 2011~2018 年 $\text{PT} \geq 50$ 的强冷湿事件共 16 d 样本作为独立预报样本，进行粒子群—模糊神经网络强冷湿事件定量集成预报试验。在强冷湿事件定量集成预报模型的输入上，采用与第 3 节相同的方法提取预报因子，经过相关计算得到，对强冷湿事件有显著影响的物理量预报因子共 41 个，其中，高度场因子 8 个，温度场因子 9 个，湿度场因子 7 个，风场（经向风和纬向风）因子 11 个，垂直速度场因子 6 个；41 个预报因子以组合因子为主，因子相关系数的绝对值 ($|R|$) 范围在 [0.30, 0.58] 之间（达到 0.001 相关显著性水平，具体因子表略）。本文以这 41 个因子作为模型输入，采用粒子群—模糊神经网络定量集成预报方法，进行影响广西的持续性冷湿极端天气的定量预报试验。

由于模糊神经网络本身并不提供如何构造学习矩阵的方法，为了控制网络结构规模，提高粒子群—模糊神经网络集成预报模型的泛化能力和有利于进行实际预报应用，本文采用在大气学科领域内广泛应用的主成分分析方法 (Principal Component Analysis, 简称 PCA) 建立合理的粒子群—模糊神经网络学习矩阵 (施能, 2008)。虽然多变量样本能够提供更丰富的预报信息，但是变量之间存在的复共线性关系会影响模糊神经网络模型的预报精度，而 PCA 方法可以从具有随机噪声干扰的气象场提

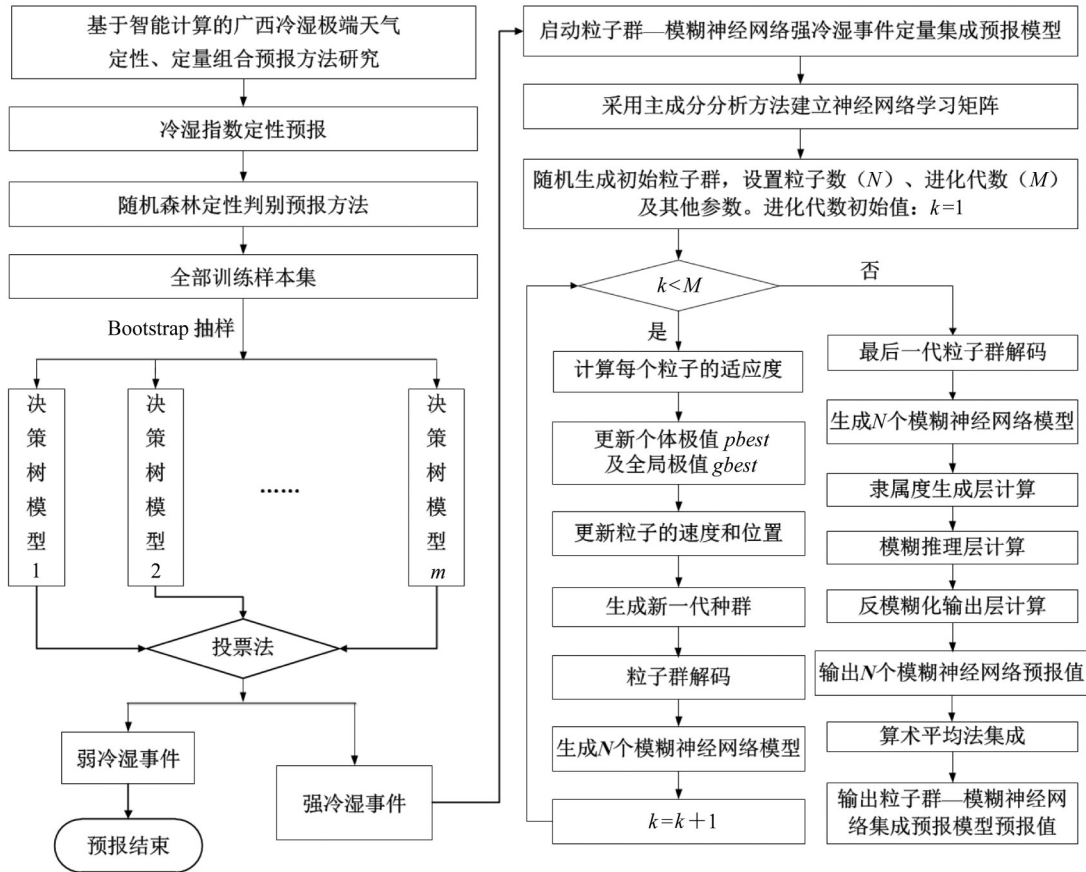


图4 广西冷湿极端天气的随机森林算法的定性预报与粒子群—模糊神经网络集成定量组合预报的建模计算流程

Fig. 4 Flowchart of the combination prediction model for cold-wet extreme weather in Guangxi employing both a qualitative forecast based on a random forest algorithm and a quantitative forecast using the Particle Swarm Optimization-Fuzzy Neural Network ensemble model

取主要信号特征，得到相互正交的主成分，不会产生复共线性影响。因此，本文采用PCA方法将41个预报因子压缩到少数几个主要模态的特征，排除随机干扰。具体做法如下：

以1956~2017年每年12月至次年2月研究区域内的极端冷湿指数天气过程为预报对象，取建模样本为1956~2010年 $PT \geq 50$ 强冷湿事件共115 d，并以2011~2018年 $PT \geq 50$ 强冷湿天气过程的16 d作为独立预报样本，预报因子为41个。首先对第1个独立预报样本进行预报，取115个建模样本加上第1个独立样本的41个物理量预报因子作PCA计算，得到41样本长度为116的主成分。在此以累积解释方差百分率达80%左右作为选取主成分个数的标准，这些方差贡献大的入选因子能更好地反映全部因子的预报信息，从而预报效果也会更好(Kidson, 1975)。经计算，41个主成分因子的前9个主成分的累积方差贡献达到了81.04% (见表1)。

因此，取这前9个主成分作为第1个独立预报样本的预报模型输入，采取同样的入选原则计算第2~第16个独立预报样本的主成分。

本文在进行24 h强冷湿指数极端天气的定量预报试验时，集成预报成员的生成是采用第4.1节介绍的粒子群—模糊神经网络集成个体生成方法，以强过程冷湿指数事件($PT \geq 50$)选取的41个预报因子作PCA提取主成分作为模型输入，建立相应的粒子群—模型神经网络集成预报模型。在粒子群算法的进化计算过程中：进化代数为30；粒子数为40 (即模糊神经网络预报个体数)；为了使粒子的个体经验和群体经验具有相同的影响力，设置 $c_1 = c_2 = 2$ ；惯性权重的最小值和最大值分别为0.4和0.9；输入节点为模型输入的主成分个数；输出节点为1；模糊神经网络模糊规则数的搜索空间为输入节点的0.2~2.0倍；模糊神经网络的模糊化中心、宽度和输出层的连接权值搜索空间为[0, 2]；

表1 24小时冷湿指数极端天气定量预报第1个独立样本入选的主成分方差贡献及累积方差贡献

Table 1 Variance contribution and accumulative variance contribution of the principal components selected for the 1st independent prediction sample in the 24-h cold-wet extreme weather quantitative forecast

主成分排序	1	2	3	4	5	6	7	8	9
方差贡献	44.57%	10.00%	6.35%	4.72%	4.46%	3.23%	2.77%	2.56%	2.39%
累积方差贡献	44.57%	54.56%	60.91%	65.63%	70.09%	73.32%	76.09%	78.65%	81.04%

模糊神经网络的学习因子为0.8。进化计算结束后，对粒子群的40个个体解码，得到40个模糊神经网络预报个体。对40个集成预报个体的预报值，再运用算术平均法进行集成，建立最终的定量集成预报模型。

在定量集成预报模型对16个独立预报样本进行预报试验时，先用115个建模样本建立预报模型对第1个独立样本作预报，然后，再以115个建模样本加上前面第1个独立样本，以116个样本作为建模样本，对第2个独立样本进行预报……，以此类推，最后是用115加前15个独立样本即130个样本作为建模样本对最后一个（第16个）独立样本作预报，这样可以使独立样本的预报与实际预报一致。因为在实际业务预报中，利用已知的前期预报因子和预报量即可建立预报模型进行预报。逐次预报过程中，粒子群算法和模糊神经网络的所有参数保持不变，每次预报都对建模样本重新进行PCA计算，按照相同的标准提取主成分作为模型输入。表2为采用粒子群—模糊神经网络集成定量预报模型，对24h广西冷湿极端天气独立预报样本的预报误差情况。经统计得到，16个独立预报样本的预报平均绝对误差为16.92，平均相对误差为23.02%，均方根误差为20.92。

5 定量预报的对比分析研究

为了比较本文运用粒子群—模糊神经网络集成定量预报建模方法对广西冷湿极端天气过程的预报能力，采用逐步回归方法，通过控制 F 值得到不同的因子组合，以相同的建模样本和独立预报样本，进行广西冷湿指数极端天气非线性预报方法与线性预报方法的试验对比分析。并进一步利用与集成模型相同的主成分作为模型输入，运用普通的模糊神经网络方法对同样的广西冷湿指数极端天气样本个例进行预报试验，考察粒子群—模糊神经网络集成模型的预报性能与一般模糊神经网络预报模型的差异。

表2 粒子群—模糊神经网络(PSO-FNN)集成预报模型与逐步回归预报方法对广西冷湿指数极端天气独立预报样本的误差统计

Table 2 Prediction error statistics for wet-cold extreme weather using the Particle Swarm Optimization-Fuzzy Neural Network (PSO-FNN) ensemble model and stepwise regression method

预报模型	因子数	平均绝对误差	平均相对误差	均方根误差
PSO-FNN集成	9~10	16.92	23.02%	20.92
逐步回归($F=2$)	7~10	33.80	44.17%	38.28
逐步回归($F=3$)	7~9	33.33	43.71%	38.44
逐步回归($F=4$)	7~8	31.14	40.55%	34.80
逐步回归($F=5$)	7	33.01	43.30%	37.51
逐步回归($F=6$)	5~7	33.82	43.64%	38.16

5.1 与逐步回归方法的对比分析

逐步回归方法是一种比较成熟的统计方法，没有可调参数便于客观比较。本文以24h广西冷湿指数极端天气定量预报结果作为比较对象，将这种新的粒子群—模糊神经网络集成预报方法与相应的逐步回归方法的预报精度作对比分析。

为保证对比的合理性，对初选得到的41个物理量预报因子，选取与粒子群—模糊神经网络集成预报试验中相同或相近的主成分因子数（大约为9~10个），建立24h广西冷湿指数极端天气的逐步回归预报方程，并以相同的建模样本和独立预报样本，对16个独立样本进行预报试验（见表2）。由于不同的预报因子组合会得到不同的预报方程，这些预报方程的预报能力是有差异的。因此，为了进一步对比不同 F 值建立的逐步回归模型对广西冷湿指数极端天气的预报误差情况，本文取 $F=2\sim 6$ ，以相同的样本个例和逐次预报方法，分别计算了不同的逐步回归方程对冷湿指数极端天气的预报能力，预报误差情况见表2。

由表2的对比分析可以看出，虽然两种预报模型的建模样本和独立样本完全相同，但是粒子群—

模糊神经网络集成预报模型的预报能力比逐步回归方法显著提高，其中，集成预报模型对 24 h 广西冷湿指数极端天气的预报平均绝对误差比逐步回归方法预报模型下降了 49.94% ($F=2$)、49.23% ($F=3$)、45.66% ($F=4$)、48.74% ($F=5$) 及 49.97% ($F=6$)。图 5 给出了广西冷湿指数极端天气独立预报样本实况与粒子群—模糊神经网络集成预报模型、预报效果较好的逐步回归方法 ($F=4$) 预报值对比图。分析图 5 可知，独立预报样本均为实况 $PT \geq 50$ 的样本个例，集成模型的预报值均在 50 以上，而逐步回归方法有 2 个预报样本的预报值小于 50，分别为 30.16、49.69，对应的集成模型预报值分别为 65.93 (实况 84.16)、71.23 (实况 74.98)。特别地，对于 $PT \geq 100$ 的 4 个冷湿极端天气样本，集成模型的预报正确 3 个样本，而逐步回归方法仅报出 1 个 $PT \geq 100$ 样本。因此，集成预报模型对 $PT \geq 50$ 的广西强冷湿极端天气定量预报具有良好的表现。

为了更进一步客观地比较非线性的粒子群—模糊神经网络集成模型与线性的逐步回归方法对广西冷湿指数极端天气的预报性能差异，以初选得到的 41 个物理量预报因子为基础，利用与集成模型相同的主成分作为逐步回归模型的输入，建立了基于主成分分析的多元回归预报模型，对 24 h 广西冷湿指数极端天气进行预报试验 (见表 3)。由表 3 的统计结果可以看到，在主成分预报因子，建模样本及独立预报样本完全一致的情况下，以主成分作为模型输入建立的多元回归方法，其预报误差明显大于粒子群—模糊神经网络集成模型。

分析表 2 和表 3 可知，无论 F 取何值，逐步回归方法的独立样本预报绝对误差、相对误差和均方

根误差均明显大于粒子群—模糊神经网络集成方法的结果。这主要是因为多元回归预报方法是采用线性回归分析方法建立的预报方程，而粒子群—模糊神经网络集成方法是一种智能计算的非线性预报建模方法，具有很强的非线性映射和自适应学习能力，可以更好地反映冷湿指数极端天气事件的无序性、非线性变化特点。同时，集成模型采用进化计算的粒子群优化算法，通过较强的全局搜索能力，在期望输出相同的情况下构造了众多的模糊神经网络优良个体，由这些个体作集合，作为集成模型的输出结论，因而，集成模型的预报精度明显提高且预报效果稳定，独立预报样本的预报值比采用多元回归方法更接近实况。

5.2 与普通模糊神经网络进行预报试验的对比分析

为了客观考察集成模型的预报性能是否优于普通的单个模糊神经网络预报模型，以粒子群—模糊神经网络集成模型试验中相同的主成分作为模型输入，同样采取逐次预报的方式，运用普通模糊神经网络方法对相同的广西冷湿指数极端天气建模样本和独立预报样本进行预报计算分析。对于一般的模

表 3 PSO-FNN 集成预报模型与基于 PCA 的多元回归预报模型对广西冷湿指数极端天气独立预报样本的误差统计

Table 3 Prediction error statistics for wet-cold extreme weather forecasts in Guangxi using the PSO-FNN ensemble model and multiple regression method based on the PCA (Principal Component Analysis)

预报模型	平均绝对误差	平均相对误差	均方根误差
PSO-FNN 集成	16.92	23.02%	20.92
多元回归-PCA	22.71	32.00%	25.93

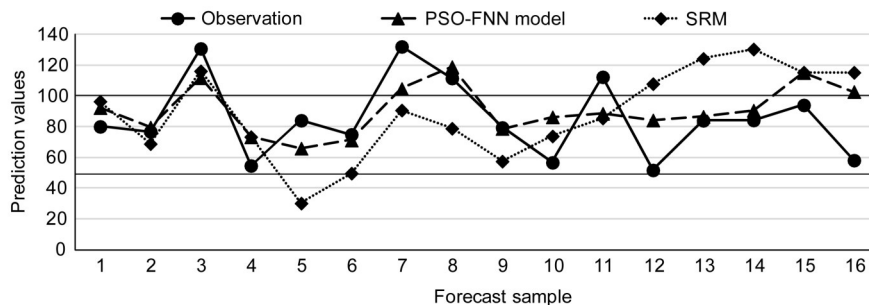


图 5 广西冷湿指数极端天气 16 个独立预报样本实况 (圆点) 与粒子群—模糊神经网络 (PSO-FNN) 集成预报模型 (三角形)、逐步回归方法 ($F=2$, 菱形) 的预报值对比

Fig. 5 Predicting values of 16 independent samples of wet-cold extreme weather forecasts in Guangxi using the PSO-FNN ensemble model (triangle), stepwise regression method (SRM) ($F=2$, diamond), and corresponding observation values (solid circle)

神经网络模型，由于没有粒子群优化算法客观确定参数和网络结构，所以模糊化中心、宽度和反模糊化输出层的连接权值均由随机数生成，模糊规则数根据经验取值，采用误差反向传播（BP）算法确定网络结构，而为了客观对比分析，其余参数设置及其计算过程也完全与粒子群—模糊神经网络集成模型的一致，预报结果见表4。

由表4的统计结果可以看到，在PCA预报因子，预报建模样本及独立预报样本完全一致的情况下，基于普通模糊神经网络的广西冷湿指数极端天气的预报平均绝对误差均高于本文集成模型所预报的独立预报样本的结果。进一步对独立样本的预报误差作统计发现，集成模型独立样本的预报平均绝对误差比普通模糊神经网络模型预报结果下降了14.37%。主要是因为新方案中改进了一般普通单个模糊神经网络采用BP算法确定网络结构的缺陷，所运用的粒子群算法是仿照集合天气的数值预报思想，将粒子群与模糊神经网络相结合，利用粒子群算法的全局搜索能力优化模糊神经网络的连接权值和隶属度函数等参数，充分发挥各算法优点，

表4 PSO-FNN集成预报模型与普通神经网络（FNN）预报模型对广西冷湿指数极端天气独立预报样本的误差统计

Table 4 Prediction error statistics for wet-cold extreme weather forecasts in Guangxi using the PSO-FNN ensemble model and normal Fuzzy Neural Network (FNN)

预报模型	平均绝对误差	平均相对误差	均方根误差
PSO-FNN集成	16.92	23.02%	20.92
FNN	19.76	24.56%	23.95

通过改进、提高集成个体的预报能力和集成个体种群差异性，提高了集成预报模型的预报精度。另外，在普通模糊神经网络中，模糊规则数是根据经验随机设定，并且在逐次计算过程中保持不变，因而对网络计算结果有较大的影响。而在集成模型中，模糊规则数由粒子群算法优化计算确定，随着适应度值进行调整，因此，集成模型网络结构稳定，预报结果可靠，预报效果好于单个的普通模糊神经网络。

5.3 预报技巧水平

此外，为了进一步考察这种新的粒子群—模糊神经网络集成预报模型建模方法，对实际业务预报是否具有预报技巧，再利用预报技巧水平计算公式（黄颖等，2014）计算了本文所提出的这种新方案相对于预报效果较好的普通模糊神经网络及基于主成分分析的多元回归模型、常规的逐步回归方法（ $F=2\sim 6$ ）的预报技巧水平分别为：14.37%，25.50%，49.94%（ $F=2$ ）、49.23%（ $F=3$ ）、45.66%（ $F=4$ ）、48.74%（ $F=5$ ）及49.97%（ $F=6$ ），均为正技巧水平（见图6）。分析发现，16个独立样本中，PT值大于100的有4个，线性建模方法对这端天气事件的预报能力一般比较弱，而本文运用粒子群—模糊神经网络方法处理这种极端天气事件（非线性）的能力相对较好，因而能取得更好的预报效果。与单个的普通模糊神经网络模型相比，粒子群—模糊神经网络集成预报模型是由多个个体差异的模糊神经网络结果合成，因而预报精度高于单个模型。分析图6的预报技巧可知，粒子群—模糊神经网络集成预报模型的广西冷湿指数极端

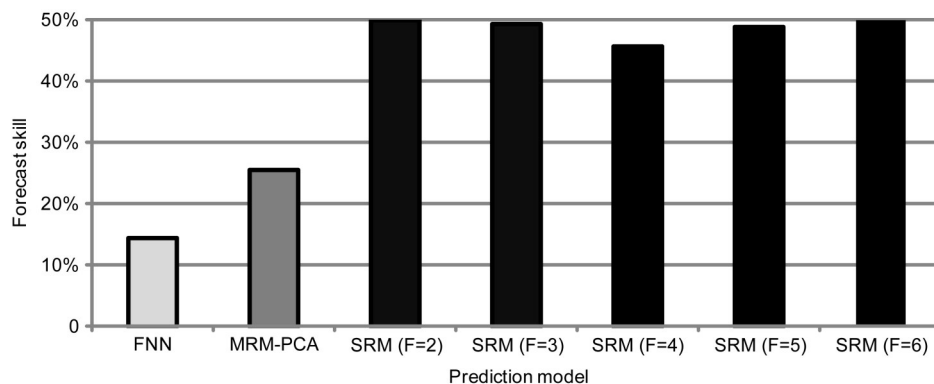


图6 基于粒子群—模糊神经网络方法的广西冷湿极端天气集成预报模型相对于普通模糊神经网络（FNN）预报模型、基于PCA的多元回归方法及逐步回归方法的预报技巧水平

Fig. 6 Forecast skill of the PSO-FNN ensemble prediction model relative to the normal FNN, the multiple regression model based on PCA (MRM-PCA), and the stepwise regression method (SRM) for wet-cold extreme weather prediction

天气预报的预报技巧相对于单个神经网络预报模型及多元回归方法、逐步回归方法均为正的预报技巧水平,说明新的模型具备更好的预报能力,而且本文在建立粒子群—模糊神经网络的集成预报模型时,所有参数都取固定,从而使预报具有很好的客观性。

6 小结

本文尝试利用表征低温雨雪冰冻过程强弱程度的冷湿指数作为预报量,通过分析提取物理意义明确的物理量预报因子作为模型输入,分别采用随机森林算法建立的冷湿指数定性预报和利用粒子群—模糊神经网络构建的集成定量预报,建立了基于智能计算的冷湿极端天气定性、定量组合预报方法,对影响广西的强冷湿事件进行预报试验。结果表明:

(1) 由于随机森林算法不需要考虑一般回归分析面临的多元共线性的问题,本文以相关普查得到的预报因子全部作为模型输入,利用随机森林分类方法可以随机选择决策树节点划分特征的能力,建立了基于随机森林的冷湿极端天气定性判别预报模型。预报试验结果表明,该定性判别预报模型计算速度快,准确识别出所有的冷湿极端天气事件,具有较高的预报准确率。

(2) 在上述定性判别预报的基础上,进一步仿照数值天气预报模式的集合预报思想,建立了基于粒子群—模糊神经网络集成方法的广西强冷湿指数极端天气的定量预报模型。该模型设计了以粒子群算法为基础的模糊神经网络集成个体生成技术方法,通过改进、提高集成个体的预报能力和种群差异性,提高集成预报模型的预报精度,因此,该集成预报模型的泛化性能显著提高,具有较好的实用性和推广价值。

(3) 在建模样本和独立预报样本完全相同的情况下,粒子群—模糊神经网络集成预报模型的预报精度不仅比线性的逐步回归方法有显著提高,而且比采用相同的模型输入作为预报因子分别进行逐步回归和普通模糊神经网络模型的预报结果也有明显改进。另外,通过计算表明,新模型相对逐步回归方法和单个的普通模糊神经网络的预报均为正的预报技巧水平,说明该方法对小概率极端天气有更好的预报能力。

(4) 利用不同智能计算方法构造的定性、定量的预报模型和组合预报建模方法,比较符合极端天气小概率事件的预报特点,是一种对极端天气预报建模方法的新尝试。并且这种从定性到定量的逐步推进的预报分析方法,具有比较容易理解的分析依据,值得进一步深入研究。在今后的实际预报应用中,可以尝试采用不同的数值预报产品资料构建不同的预报模型,并进行模糊神经网络集成预报个体差异和性能的对比试验分析,进一步增强个体差异度,提高粒子群—模糊神经网络集成预报模型的泛化性能。而先定性预报再进一步作定量预报,这两种预报结果趋势上的一致性,也增强了最终预报结果的可靠性。

参考文献 (References)

- Barszcz Agnieszka, Milbrandt Jason A., Thériault Julie M. 2018. Improving the explicit prediction of freezing rain in a kilometer-scale numerical weather prediction model [J]. *Weather and Forecasting*, 33(3): 767–782. doi:10.1175/WAF-D-17-0136.1
- Bao Qing, Yang Jing, Liu Yimin, et al. 2010. Roles of anomalous Tibetan Plateau warming on the severe 2008 winter storm in central-southern China [J]. *Monthly Weather Review*, 138(6): 2375–2384. doi:10.1175/2009MWR2950.1
- Breiman Leo. 2001. Random forests [J]. *Machine Learning*, 45(1): 5–32. doi:10.1023/A:1010933404324
- 布和朝鲁, 彭京备, 谢作威, 等. 2018. 冬季大范围持续性极端低温事件与欧亚大陆大型斜脊斜槽系统研究进展 [J]. *大气科学*, 42(3): 656–676.
- Bueh Cholaw, Peng Jingbei, Xie Zuwei, et al. 2018. Recent progresses on the studies of wintertime extensive and persistent extreme cold events in China and large-scale tilted ridges and troughs over the Eurasian Continent [J]. *Chinese Journal of Atmospheric Sciences (in Chinese)*, 42 (3): 656–676. doi:10.3878/j.issn.1006-9895.1712.17249
- 丁一汇, 王遵娅, 宋亚芳, 等. 2008. 中国南方2008年1月罕见低温雨雪冰冻灾害发生的原因及其与气候变暖的关系 [J]. *气象学报*, 66 (5): 809–825.
- Ding Yihui, Wang Zunya, Song Yafang, et al. 2008. Causes of the unprecedented freezing disaster in January 2008 and its possible association with the global warming [J]. *Acta Meteorologica Sinica (in Chinese)*, 66(5): 808–825. doi:10.3321/j.issn:0577-6619.2008.05.014
- 杜小玲, 高守亭, 彭芳. 2014. 2011年初贵州持续低温雨雪冰冻天气成因研究 [J]. *大气科学*, 38(1): 61–72.
- Du Xiaoling, Gao Shouting, Peng Fang. 2014. Study of the 2011 freezing rain and snow storm in Guizhou [J]. *Chinese Journal of Atmospheric Sciences (in Chinese)*, 38 (1): 61–72. doi:10.3878/j.issn.1006-9895.2013.12119

- 龚志强, 王晓娟, 崔冬林, 等. 2012. 区域性极端低温事件的识别及其变化特征 [J]. 应用气象学报, 23(2): 195-204. Gong Zhiqiang, Wang Xiaojuan, Cui Donglin, et al. 2012. The Identification and changing characteristics of regional low temperature extreme events [J]. Journal of Applied Meteorological Science (in Chinese), 23(2): 195-204. doi:10.11898/1001-7313.20120208
- 何伟, 李亚伟, 金栋, 等. 2004. 基于PSO的模糊人工神经网络径流预报模型 [J]. 气象水文海洋仪器, 2: 18-21. He Wei, Li Yawei, Jin Dong, et al. 2004. Fuzzy artificial neural network runoff prediction model based on PSO [J]. Meteorological, Hydrological and Marine Instrument (in Chinese), 2: 18-21. doi: 10.3969/j.issn.1006-009X.2004.02.005
- Herman G R, Schumacher R S. 2018. "Dendrology" in numerical weather prediction: What random forests and logistic regression tell us about forecasting extreme precipitation [J]. Monthly Weather Review. 146(6): 1785-1812. doi:10.1175/MWR-D-17-0307.1
- 黄颖, 金龙, 黄小燕, 等. 2014. 基于局部线性嵌入的人工智能台风强度集合预报模型 [J]. 气象, 40(7): 806-815. Huang Ying, Jin Long, Huang Xiaoyan, et al. 2014. An artificial intelligence ensemble prediction scheme for typhoon intensity using the locally linear embedding [J]. Meteorological Monthly (in Chinese), 40(7): 806-815. doi:10.7519/j.issn.1000-0526.2014.07.004
- Huang Ying, Jin Long, Zhao Huasheng, et al. 2018. Fuzzy neural network and LLE algorithm for forecasting precipitation in tropical cyclones: Comparisons with interpolation method by ECMWF and stepwise regression method [J]. Natural Hazards, 91(1): 201-220. doi:10.1007/s11069-017-3122-x
- 江漫, 于甜甜, 钱维宏. 2014. 我国南方冬季低温雨雪冰冻事件的大气扰动信号分析 [J]. 大气科学, 38(4): 813-824. Jiang Man, Yu Tiantian, Qian Weihong. 2014. Analysis on atmospheric anomalous signals of winter low temperature and snow ice storms in southern China [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 38(4): 813-824. doi: 10.3878/j.issn.1006-9895.2013.13217
- 金龙. 2004. 神经网络气象预报建模理论与应用 [M]. 北京: 气象出版社, 75-185. Jin Long. 2004. Modeling Theory Method and Application for Weather Forecast Based on Neural Network (in Chinese) [M]. Beijing: China Meteorological Press, 75-185
- 金龙, 林熙, 金健, 等. 2003. 模块化模糊神经网络的数值预报产品释用预报研究 [J]. 气象学报, 61(1): 78-84. Jin Long, Lin Xi, Jin Jian, et al. 2003. Study on interpretation prediction of numerical weather prediction products based on modular fuzzy neural network [J]. Acta Meteorologica Sinica (in Chinese), 61(1): 78-84. doi: 10.3321/j.issn:0577-6619.2003.01.008
- Jin Long, Yao Cai, Huang Xiaoyan. 2008. A nonlinear artificial intelligence ensemble prediction model for typhoon intensity [J]. Monthly Weather Review, 136(12): 4541-4554. doi: 10.1175/2008mwr2269.1
- Jin Long, Zhu Jiashun, Huang Ying, et al. 2015. A nonlinear statistical ensemble model for short-range rainfall prediction [J]. Theoretical and Applied Climatology, 119(3-4): 791-807. doi: 10.1007/s00704-014-1161-8
- Kalnay E, Kanamitsu M, Kistler R, et al. 1996. The NCEP/NCAR 40-year reanalysis project [J]. B. Am. Meteorol. Soc., 77(3): 437-472. doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2
- Kennedy J, Russell E. 1995. Particle swarm optimization [C]// Proceedings of the IEEE International Conference on Neural Networks, IEEE Service Center, Piscataway, NJ, 1942-1948. doi:10.1109/ICNN.1995.488968
- Kidson J W. 1975. Eigenvector analysis of monthly mean surface data [J]. Monthly Weather Review, 103(3): 177-186. doi.org/10.1175/1520-0493(1975)103<0177:EAOMMS>2.0.CO;2
- 李崇银, 杨辉, 顾薇. 2008. 中国南方雨雪冰冻异常天气原因的分析 [J]. 气候与环境研究. 13(2): 113-122. Li Chongyin, Yang Hui, Gu Wei. 2008. Cause of severe weather with cold air, freezing rain and snow over South China in January 2008 [J]. Climatic and Environmental Research (in Chinese), 13(2): 113-122. doi:10.3878/j.issn.1006-9585.2008.02.01
- 李建更, 高志坤. 2009. 随机森林: 一种重要的肿瘤特征基因选择法 [J]. 生物物理学报, 25(1): 51-56. Li Jiange, Gao Zhikun. 2009. Random forests: An important feature genes selection method of tumor [J]. Acta Biophysica Sinica (in Chinese), 25(1): 51-56.
- 李欣海. 2013. 随机森林模型在分类与回归分析中的应用 [J]. 应用昆虫学报, 50(4): 1190-1197. Li Xinhai. 2013. Using "random forest" for classification and regression [J]. Chinese Journal of Applied Entomology (in Chinese), 50(4): 1190-1197. doi: 10.7679/j.issn.2095-1353.2013.163
- 刘吉敏, 黄泓, 吐莉尼沙. 2018. 随机森林模型判定热带气旋变性完成时间 [J]. 气象科技. 46(2): 307-315. Liu Jimin, Huang Hong, Tulinisha. 2018. Applying random forests to determine completion time of extratropical transitioning tropical cyclones [J]. Meteorological Science and Technology (in Chinese), 46(2): 307-315. doi:10.19517/j.1671-6345.20170246
- 陆虹, 翟盘茂, 覃卫坚, 等. 2015. 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, 26(5): 513-524. Lu Hong, Zhai Panmao, Qin Weijian, et al. 2015. A particle swarm optimization neural network ensemble prediction model for persistent freezing rain and snow storm in southern China [J]. Journal of Applied Meteorological Science (in Chinese), 26(5): 513-524. doi:10.11898/1001-7313.20150501
- 彭艳, 王钊, 刘安麟, 等. 2010. 2008年1月中国南部低温雨雪冰冻天气特征及其与东亚大气环流异常探讨 [J]. 大气科学学报, 33(5): 634-640. Peng Yan, Wang Zhao, Liu Anlin, et al. 2010. Characteristics of the freezing rain/heavy snow process in South China and the anomalies of atmospheric circulation over East Asia in January 2008 [J]. Transactions of Atmospheric Sciences (in Chinese), 33(5): 634-640. doi:10.3969/j.issn.1674-7097.2010.05.016
- Ren Fumin, Cui Donglin, Gong Zhiqiang, et al. 2012. An objective

- identification technique for regional extreme events [J]. *Journal of Climate*, 25(20): 7015–7027. doi:10.1175/JCLI-D-11-00489.1
- Ressler G M, Milrad S M, Atallah E H, et al. 2012. Synoptic-scale analysis of freezing rain events in Montreal, Quebec, Canada [J]. *Weather and Forecasting*, 27(2): 362–378. doi: 10.1175/WAF-D-11-00071.1
- 师彪, 李郁侠, 于新花, 等. 2010. 基于改进粒子群—模糊神经网络的短期电力负荷预测 [J]. *系统工程理论与实践*, 30(1): 157–166.
- Shi Biao, Li Yuxia, Yu Xinhua, et al. 2010. Short-term load forecasting based on modified particle swarm optimizer and fuzzy neural network model [J]. *Systems Engineering Theory & Practice* (in Chinese), 30(1): 157–166.
- 施能. 2008. 气象科研与预报中的多元分析方法 [M]. 北京: 气象出版社, 114–122. Shi Neng. 2008. *Multivariate Analysis Method in Meteorological Research and Forecasting* (in Chinese) [M]. Beijing: China Meteorological Press, 114–122.
- Shi Yuhui, Russell Eberhart. 1998. A modified particle swarm optimizer [C]. *The 1998 IEEE International Conference on Evolutionary Computation Proceedings*. Anchorage, AK, USA, 4–9 May 1998, 69–73. doi: 10.1109/ICEC.1998.699146
- 陶诗言, 卫捷. 2008. 2008年1月我国南方严重冰雪灾害过程分析 [J]. *气候与环境研究*, 13(4): 337–350. Tso Shiyan, Wei Jie. 2008. Severe snow and freezing-rain in January 2008 in the southern China [J]. *Climatic and Environmental Research* (in Chinese), 13(4): 337–350. doi:10.3878/j.issn.1006-9585.2008.04.01
- Tobin D M, Kumjian M R. 2017. Polarimetric radar and surface-based precipitation-type observations of ice pellet to freezing rain transitions [J]. *Weather and Forecasting*, 32(6): 2065–2082. doi: 10.1175/WAF-D-17-0054.1
- 钱维宏, 张宗婕. 2012. 南方持续低温冻雨事件预测的前期信号 [J]. *地球物理学报*, 55(5): 1501–1512. Qian Weihong, Zhang Zongjie. 2012. Precursors to predict low-temperature freezing-rain events in southern China [J]. *Journal of Geophysics* (in Chinese), 55(5): 1501–1512. doi:10.6038/j.issn.0001-5733.2012.05.007
- Qian Xi, Miao Qilong, Zhai Panmao, et al. 2014. Cold-wet spells in mainland China during 1951–2011 [J]. *Natural Hazards*, 74(2): 931–946. doi:10.1007/s11069-014-1227-z
- 沈润平, 郭佳, 张婧娴. 2017. 基于随机森林的遥感干旱监测模型的构建 [J]. *地球信息科学学报*, 19(1): 125–133. Shen Runping, Guo Jia, Zhang Jingxian, et al. 2017. Construction of a drought monitoring model using the random forest based remote sensing [J]. *Journal of Geo-information Science* (in Chinese), 19(1): 125–133. doi:10.3724/SP.J.1047.2017.00125
- 万素琴, 周月华, 李兰, 等. 2008. 低温雨雪冰冻极端气候事件的多指标综合评估技术 [J]. *气象*, 34(11): 40–46. Wan Suqin, Zhou Yuehua, Li Lan, et al. 2008. A multi-index synthetic assessment method for extreme climate events of sleet and freezing with low temperature [J]. *Meteorological Monthly* (in Chinese), 34(11): 40–46. doi:10.7519/j.issn.1000-0526.2008.11.006
- 王东海, 柳崇健, 刘英, 等. 2008. 2008年1月中国南方低温雨雪冰冻天气特征及其天气动力学成因的初步分析 [J]. *气象学报*, 66(3): 405–422. Wang Donghai, Liu Congjian, Liu Ying, et al. 2008. A preliminary analysis of features and causes of the snow storm event over the southern China in January 2008 [J]. *Acta Meteorologica Sinica* (in Chinese), 66(3): 405–422. doi:10.3321/j.issn:0577-6619.2008.03.011
- 王士同. 1998. 神经模糊系统及其应用 [M]. 北京: 北京航空航天大学出版社, 126–168. Wang Shitong. 1998. *Neural Fuzzy Systems and Its Applications* (in Chinese) [M]. Beijing: Beihang University Press, 126–168.
- 王晓娟, 沈柏竹, 龚志强, 等. 2013. 中国冬季区域性极端低温事件分类及其与气候指数极端性的联系 [J]. *物理学报*, 62(22): 229201 Wang Xiaojuan, Shen Baizhu, Gong Zhiqian et al. 2013. The classification of winter regional extreme low temperature events in China and their corresponding relationship to climatic indices extreme anomaly [J]. *Acta Physica Sinica* (in Chinese), 62(22): 229201. doi:10.7498/aps.62.229201
- 衣治安, 牟春苗, 孙寅萍. 2014. 基于改进粒子群算法的模糊神经网络 [J]. *计算机系统应用*, 23(6): 135–140. Yi Zhihan, Mu Chunmiao, Sun Yinping. 2014. Fuzzy neural network based on improved particle swarm algorithms [J]. *Computer Systems and Applications* (in Chinese), 23(6): 135–140. doi:10.3969/j.issn.1003-3254.2014.06.024
- 余胜男, 陈元芳, 顾圣华, 等. 2016. 随机森林在降水量长期预报中的应用 [J]. *南水北调与水利科技*, 14(1): 78–83. Yu Shengnan, Chen Yuanfang, Gu Shenghua, et al. 2016. Long term rainfall forecasting based on random forest [J]. *South to North Water Transfers and Water Science & Technology* (in Chinese), 14(1): 78–83. doi: 10.13476/j.cnki.nsbdqk.2016.01.013
- Zhang Huqiang, Qin Jun, Li Yun. 2011. Climatic background of cold and wet winter in southern China. Part I: Observational analysis [J]. *Climate Dynamics*, 37(11–12): 2335–2354. doi:10.1007/s00382-011-1022-4
- 张晓羽, 李凤日, 甄贞, 等. 2016. 基于随机森林模型的陆地卫星-8遥感影像森林植被分类 [J]. *东北林业大学学报*, 44(6): 53–57. Zhang Xiaoyu, Li Fengri, Zhen Zhen, et al. 2016. Forest vegetation classification of landsat-8 remote sensing image based on random forests model [J]. *Journal of Northeast Forestry University* (in Chinese), 44(6): 53–57, 74.
- 赵华生, 黄小燕, 黄颖. 2018. ECMWF集合预报产品在广西暴雨预报中的释用 [J]. *应用气象学报*, 29(3): 344–353. Zhao Huasheng, Huang Xiaoyan, Huang Ying. 2018. Application of ECMWF ensemble forecast products to rainstorm forecast in Guangxi [J]. *Journal of Applied Meteorological Science* (in Chinese), 29(3): 344–353. doi:10.11898/1001-7313.20180308
- 赵思雄, 孙建华. 2008. 2008年初南方雨雪冰冻天气的环流场与多尺度特征 [J]. *气候与环境研究*, 13(4): 351–367. Zhao Sixiong, Sun Jianhua. 2008. Multi-scale systems and conceptual model on

- freezing rain and snow storm over southern China during January–February 2008 [J]. Climatic and Environmental Research (in Chinese), 13(4): 351–367. doi:10.3878/j.issn.1006-9585.2008.04.02
- 邹亮, 黄琼, 李骛, 等. 2012. 基于随机森林和富集分析的阿尔茨海默症 GWA 研究 [J]. 中国科学: 生命科学, 42(8): 639–647. Zou Liang, Huang Qiong, Li Ao, et al. 2012. A genome-wide association study of Alzheimer's disease using random forests and enrichment analysis [J]. Science China: Life Science (in Chinese), 42(8): 639–647. doi:10.1007/s11783-011-0280-z
- 周秀华, 陆虹, 金龙, 等. 2019. 适用于我国华南广西区域冬季的冷湿指数计算方法研究 [J]. 灾害学, 34(2): 31–36. Zhou Xiuhua, Lu Hong, Jin Long, et al. 2019. Calculation method of cold-wet index for the winter in Guangxi region of South China [J]. Journal of Catastrophology (in Chinese), 34(2): 31–36. doi:10.3969/j.issn.1000-811X.2019.02.007
- 周岩, 王盛, 高传善, 等. 2010. 基于改进粒子群算法的模糊神经网络及其在短时天气预报中的应用 [J]. 计算机应用与软件, 27(5): 234–237. Zhou Yan, Wang Sheng, Gao Chuanshan, et al. 2010. An improved PSO-based fuzzy neural network and its application in short-term weather forecast [J]. Computer and Applications and Software (in Chinese), 27(5): 234–237. doi: 10.3969/j.issn.1000-386X.2010.05.070