

DOI:10.13476/j.cnki.nsbdcqk.2020.0006

疏杏胜,王子茹,李福威,等.基于机器学习模型的短期降雨多模式集成预报[J].南水北调与水利科技,2020,18(1):42-50.
SHU X S,WANG Z R,LI F W,et al.Short-term rainfall multi-mode integrated forecasting based on machine learning models [J].South-to-North Water Transfers and Water Science & Technology,2020,18(1):42-50.(in Chinese)

基于机器学习模型的短期降雨多模式集成预报

疏杏胜¹,王子茹¹,李福威²,彭勇¹

(1.大连理工大学 建设工程学部,辽宁 大连 116024;
2.国电电力发展股份有限公司和禹水电开发公司,辽宁 桓仁 117201)

摘要:短期降雨预报对洪水预报和水库调度极为重要,提高短期降雨预报精度有着重要的意义。以 TIGGE 资料中心的 ECMWF、CMA 及 NCEP 三个集合预报中心发布的桓仁水库流域预报降雨数据为基础,利用 ANN、ELM 以及 SVM 模型对桓仁水库流域未来 1~3 d 降雨进行多模式集成预报,以期提高预报精度,并从绝对平均误差、均方根误差、相对误差、纳什系数、预报准确率等多个方面分析了集成预报的效果。试验结果表明,基于 SVM 和 ELM 的多模式集成预报模型预报效果均优于单一模式,基于 ANN 的集成预报模型在输入因子选择合适的情况下,其预报效果也优于单一模式,三种模型中,SVM 模型对降雨预报精度改善最为明显。说明基于机器学习模型的多模式降雨集成预报方法可行且能够提高短期预报降雨精度。

关键词:多模式;机器学习模型;短期降雨;集成预报

中图分类号:TV125 文献标志码:A 开放科学(资源服务)标识码(OSID):



Short-term rainfall multi-mode integrated forecasting based on machine learning models

SHU Xingsheng¹, WANG Ziru¹, LI Fuwei², PENG Yong¹

(1. Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian 116024, China;
2. Heyu Hydropower Development Company, Guodian Electric Power Development Co., Ltd., Huanren 117201, China)

Abstract: High accuracy of short-term rainfall forecasting is of great importance for flood forecasting and reservoir operation. It can not only improve the accuracy of flood forecasting but also make the reservoir operation more scientific and reasonable. Based on the predicted rainfall of the Huanren reservoir basin using ECMWF, CMA and NCEP in the TIGGE datasets, the artificial neural network (ANN), support vector machine (SVM) and extreme learning machine (ELM) models were developed to simulate and forecast the rainfall of Huanren reservoir basin in the next 1 to 3 days, and the effect of the forecasting results were analyzed from the aspects of mean absolute error (MAE), root mean square error (RMSE), Bias, Nash-Sutcliffe efficiency coefficient (NSE) and prediction accuracy. Results showed that the integrated forecasting models based on SVM and ELM were better than the single models, and the integrated models based on ANN were better than the single models when the input factors were selected properly. Among the three integrated models, SVM model had the most obvious improvement in rainfall forecasting accuracy, which indicated that the multi-model rainfall integrated forecasting method based on machine learning model was feasible and could improve the accuracy of short-term rainfall forecasting.

Key words: multi-mode; machine learning model; rainfall; integrated forecasting

收稿日期:2019-04-14 修回日期:2019-06-10 网络出版时间:2019-06-26

网络出版地址: <http://kns.cnki.net/kcms/detail/13.1334.TV.20190625.0842.004.html>

基金项目:国家重点研发计划(2017YFC0406005)

作者简介:疏杏胜(1991—),男,安徽池州人,在读博士,主要从事水文预报方面研究。E-mail:1192770498@mail.dlut.edu.cn

在水库防洪和兴利过程中,短期降雨预报的作用十分重要。高精度短期降雨预报不仅有助于未来洪水的准确预警,为防洪决策提供可靠的依据,而且能为水库发电决策提供科学指导,避免水库弃水或放空,保证发电效益。随着数值天气预报技术的发展,越来越多的水文工作者开始将降雨预报信息用于洪水预报^[1-4]。此外,TIGGE 资料中心的建成也使得更多用户可以获得不同模式的降雨集合预报数据。然而,这些不同预报模式的预报精度存在一定差异,如何科学选择最优预报结果成为难题^[5]。因此,多模式集成预报技术成为综合利用不同模式预报产品的一个有效途径^[6]。Cartwright^[7]、Wu 等^[8]利用多元线性回归方法建立了多模式降雨集成预报模型,并获得了较好的预报效果。Lim 等^[9]也提出了一种基于正则回归的多模式降雨集成预报方法。在国内,王海霞等^[10]利用消除偏差的集合平均方法对多个中心的预报降雨进行了集成,获得了更高准确性的预报值。赵成武等^[6]利用消除偏差、集合平均、超级集合等方法对短期强降水进行了多模式集成预报。黄威等^[11]对淮河流域及其以南地区的中期强降雨展开了集成预报和对比。尽管多模式集成预报技术在提高降雨预报精度方面已取得较多的成果,但这些方法大多集中在对线性方法的使用。随着人工智能的发展,极限学习机、人工神经网络等机器学习模型越来越多的被应用于水文预测并取得较好的效果^[12-13]。有研究表明,机器学习模型能够很好的模拟水文系统的非线性关系^[14-15]。基于以上分析,本文尝试将集成预报思想与机器学习模型相结合,以欧洲中期天气预报中心(ECMWF)、中国气象局(CMA)以及美国国家环境预报中心(NCEP)三个模式每日及前期发布的预报降雨信息为输入,对桓仁水库流域未来 1~3 d 降雨进行集成预报,探索基于机器学习模型的多模式集成预报方法在桓仁水库流域的可利用性。

1 预报模型原理

1.1 人工神经网络

人工神经网络(artificial neural networks, ANN)是一种模拟人脑结构及其功能的线性动力系统,具有自组织、自学习以及较强容错性等特点,对模糊信息或非线性关系能很好地进行处理^[16]。ANN 的形式较多,本文选用 BP(back propagation)神经网络作为 ANN 建模的基本网络。

BP 神经网络结构简单,是一种利用反向传递来修正误差的多层映射网络,具有输入层、隐含层和输出层,各层内的神经元没有连接,层间神经元通过网

络权重系数相互联系,具体算法详见文献^[12]。参数合适时,BP 神经网络能够收敛到较小的模拟误差,因此参数及结构的设计是 BP 神经网络模型是否成功的关键,本文使用试错法确定 BP 神经网络结构和参数,根据训练期模拟精度而定^[17]。

1.2 支持向量机

支持向量机(support vector machine, SVM)是由 Vapnik Vladimir 和 Corinna Cortes 等人^[18]于 1995 年提出的一种学习方法。SVM 以 VC 维理论和结构风险最小原理为基础,在有限样本中寻求模型学习能力与复杂性的最佳折衷,能有效地避免传统学习方法中过学习、维数灾、局部最优等问题^[19]。SVM 模型的数学原理详见文献^[20]。要获得 SVM 的回归方程,需要确定几个因素:惩罚系数 C 、不敏感损失系数 ϵ 、核函数及其所含参数,这些因数决定了 SVM 的模拟和预测性能。本文中,核函数类型采用试错法确定,惩罚系数 C 和核函数参数 g 采用网格搜索法进行优化^[21]。

1.3 极限学习机

极限学习机(extreme learning machine, ELM)是由 Huang 等^[22-23]2006 年提出的一种单层前馈神经网络,它克服了传统神经网络训练速度慢、容易陷入局部最优等缺点^[24]。ELM 基于经验风险最小化理论,其训练过程只需要一次迭代,与其它神经网络相比具有更好的鲁棒性和泛化能力^[25]。目前 ELM 被广泛用于各种分类和回归等问题研究,其核心是随机设置输入权值和偏差,通过最小化训练误差来优化输出层的权值,主要算法步骤详见文献^[26]。利用 ELM 构建降雨预报模型,需要确定隐含层神经元个数 k 和激活函数 $g(x)$,本文采用试错法确定激活函数类型和隐含层神经元个数,根据训练期模拟精度而定。

2 多模式降雨集成预报研究

2.1 研究流域和资料处理

桓仁水库位于浑江中游。浑江发源于长白山系龙岗山脉的老爷岭南麓,是鸭绿江右侧的最大支流,全长 432 km,已自上而下先后建成了桓仁、回龙山、太平哨等水库。首级龙头桓仁水库具有不完全年调节性能,回龙山及太平哨均为日调节水库。桓仁水库流域面积为 1.04 万 km^2 ,占整个浑江流域面积的 70%,多年平均流量为 142 m^3/s 。流域多年平均年降水量 860 mm,5—10 月降雨量占全年降雨量的 75%以上。流域内共有 10 个雨量站,分别为三岔子、八道

江、八里哨、通化、三棵树、八道沟、东村、业主沟、花甸、桓仁。

所用预报降雨资料由 TIGGE 资料中心的 ECMWF、CMA 以及 NCEP 三个预报中心提供。TIGGE 是全球交互式大集合 THORPEX Interactive Grand Global Ensemble 的简称,目前全球共有欧洲中期天气预报中心(ECMWF)、中国气象局(CMA)和美国国家大气研究中心(NCAR)三个资料交换中心,这三个资料交换中心收集来自全球十多个预报中心的集合预报产品^[27]。本文选取 ECMWF、CMA 及 NCEP 2011 年 1 月 1 日至 2017 年 12 月 31 日每天北京时间 8 点和 20 点发布的降雨预报格点数据,并根据格点平均法转化为流域面雨量,预报区域为 124°25′~126°50′E, 40°25′~42°25′N,水平分辨率为 0.25°×0.25°;实测降雨为桓仁水库流域 10 个雨量站实测降雨数据,并按照面积加权法转化为面雨量,由桓仁水库管理局提供。将桓仁流域预报降雨资料和实测降雨资料划分为训练期(2011—2014 年)和验证期(2015—2017 年)两个部分,其中训练期用于预报模型建立,验证期用于模型检验。

2.2 预报因子选择

输入因子的选择对数据驱动模型十分重要。若

输入因子过多,则输入信息冗余,模型复杂度提升,模型预报精度降低;若输入因子过少,则输入信息无法充分解释输出变量的变化机理,同样影响模型预报效果。因此,为了减小输入因子选择带来的偶然性和不确定性,本文将设置 8 种不同的输入因子组合方案作为数据驱动模型的输入,即耦合三个集合预报中心近 n 天($n=1,2,3,\dots,8$,不同的输入因子组合方案 n 取值不同)发布的预报值作为输入因子。表 1 列出了 $n=1,2,3$ 时模型输入因子的详细情况。当 $n=1$ (即方案 1),采用近 1 d 的预报降雨作为输入因子。预报未来第 1 天时,输入因子为各中心(ECMWF、NCEP 和 CMA)当天发布的预见期为 1 d 的预报降雨,共计 3 个;预报未来第 2 天时,输入因子为各中心当天发布的预见期为 2 d 的预报降雨,共计 3 个。当 $n=2$ (即方案 2),采用近 2 d 预报降雨作为输入因子。预报未来第 1 天时,输入因子为各中心当天发布的预见期为 1 d 的预报降雨以及 1 d 前发布的预见期为 2 d 的预报降雨,共计 6 个;预报未来第 2 天时,输入因子为各中心当天发布的预见期为 2 d 的预报降雨以及 1 d 前发布的预见期为 3 d 的预报降雨,共计 6 个。依此类推,当 $n=8$ 时,模型输入因子个数为 24 个。

表 1 $n=1,2,3$ 时,各方案输入因子

Tab. 1 Input factors for each scheme when n is 1,2 or 3

方案号	预报对象	输入因子名称	输入因子个数
1	未来第 1 天	$F_{E,1}^k, F_{N,1}^k, F_{C,1}^k$	3
	未来第 2 天	$F_{E,2}^k, F_{N,2}^k, F_{C,2}^k$	3
	未来第 3 天	$F_{E,3}^k, F_{N,3}^k, F_{C,3}^k$	3
2	未来第 1 天	$F_{E,2}^k, F_{N,2}^k, F_{C,2}^k, F_{E,1}^k, F_{N,1}^k, F_{C,1}^k$	6
	未来第 2 天	$F_{E,3}^k, F_{N,3}^k, F_{C,3}^k, F_{E,2}^k, F_{N,2}^k, F_{C,2}^k$	6
	未来第 3 天	$F_{E,4}^k, F_{N,4}^k, F_{C,4}^k, F_{E,3}^k, F_{N,3}^k, F_{C,3}^k$	6
3	未来第 1 天	$F_{E,3}^k, F_{N,3}^k, F_{C,3}^k, F_{E,2}^k, F_{N,2}^k, F_{C,2}^k, F_{E,1}^k, F_{N,1}^k, F_{C,1}^k$	9
	未来第 2 天	$F_{E,4}^k, F_{N,4}^k, F_{C,4}^k, F_{E,3}^k, F_{N,3}^k, F_{C,3}^k, F_{E,2}^k, F_{N,2}^k, F_{C,2}^k$	9
	未来第 3 天	$F_{E,5}^k, F_{N,5}^k, F_{C,5}^k, F_{E,4}^k, F_{N,4}^k, F_{C,4}^k, F_{E,3}^k, F_{N,3}^k, F_{C,3}^k$	9

注:表中 $F_{k,j}^k$ 为预报降雨值; k 表示不同预报中心; $k=E$ 表示 ECMWF; $k=N$ 表示 NCEP; $k=C$ 表示 CMA; i 表示发布时间; $i=0$ 表示当天发布; $i=1$ 表示前 1 d 发布; j 表示预见期; $j=1$ 表示预见期为 1 d; $j=2$ 表示预见期为 2 d。

2.3 模型评价指标

为了定量评估模型的预报精度,本文采用绝对平均误差(mean absolute error, MAE)、均方根误差(root mean square error, RMSE)、相对误差(Bias)和纳什系数(nash-sutcliffe efficiency coefficient, NSE)评价模型计算结果与实测值之间的误差以及一致性,计算公式为

$$MAE = \frac{1}{N} \sum_{i=1}^N |sim_i - obs_i| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (obs_i - sim_i)^2} \quad (2)$$

$$Bias = \frac{\sum_{i=1}^N |sim_i - obs_i|}{\sum_{i=1}^N obs_i} \times 100\% \quad (3)$$

$$NSE = \frac{\sum_{i=1}^N (obs_i - sim_i)^2}{\sum_{i=1}^N (obs_i - \overline{obs})^2} \quad (4)$$

式中: obs_i 为第 i 时段的实测值; sim_i 为第 i 时段的

模型预报值; $\overline{\text{obs}}$ 为整个系列中实测值的平均值; N 为数据样本数量。MAE、RMSE 及 Bias 越小,模型预报误差越小;NSE 越接近 1,模拟效果越好。

为了进一步评价降多模式降雨集成预报方法的预报效果,本文对集成前后不同量级降雨预报的准确率进行了比较,预报准确率的计算公式为

$$\eta = \frac{m}{t} \times 100\% \quad (5)$$

式中: t 为发布预报总次数; m 为实际值位于预报等级区间内的次数,预报降雨等级分为 I、II、III、IV 共四级,各级降雨区间分别为: $[0, 0.1)$ 、 $[0.1, 10.1)$ 、

$[10.1, 25.1)$ 、 ≥ 25.1 mm。

2.4 模型建立与结果分析

以不同输入因子组合作为模型输入,以被预报日的实测降雨 P 作为输出,建立多模式降雨集成预报模型;对未来(验证期)从 TIGGE 资料中心获取的不同模式预报降雨数据,将其作为输入因子驱动模型,实现对未来降雨的预报。

按照上述方法,本文建立了 ANN、ELM 以及 SVM 三种模型,对未来 1~3 d 降雨进行预报,并对各集成模型的预报效果进行分析,结果见图 1 至图 3。

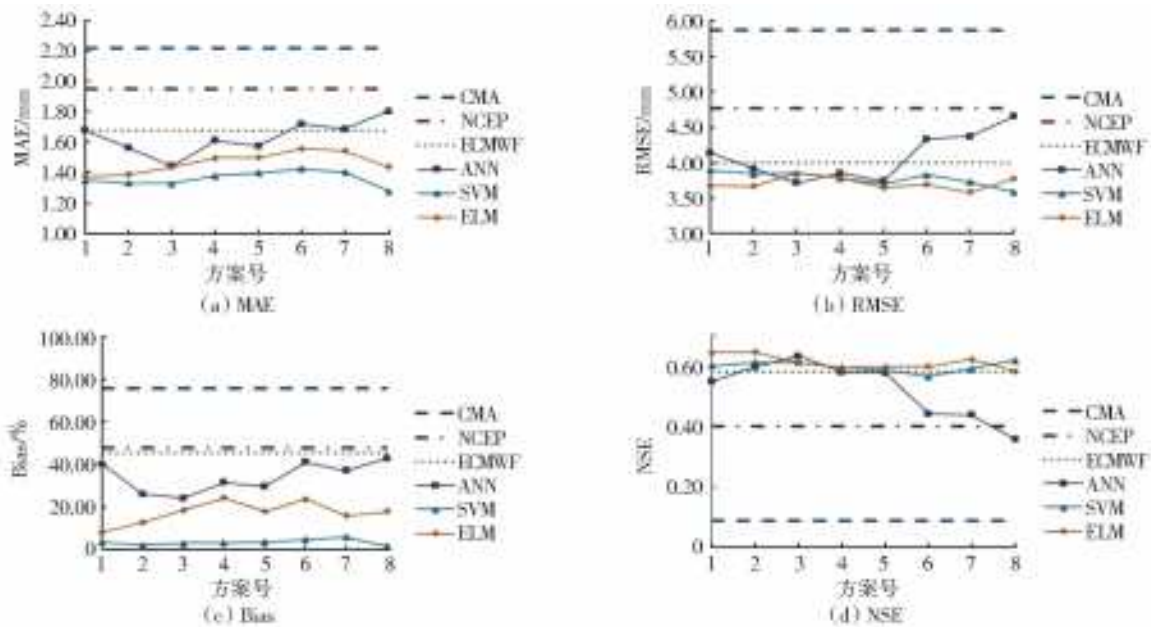


图 1 单一模式和集成预报模型验证期未来第 1 天降雨预报精度对比

Fig. 1 Comparison of the forecast accuracy of rainfall for the next day between single ensemble forecast center and three integrated forecasting schemes in verification period

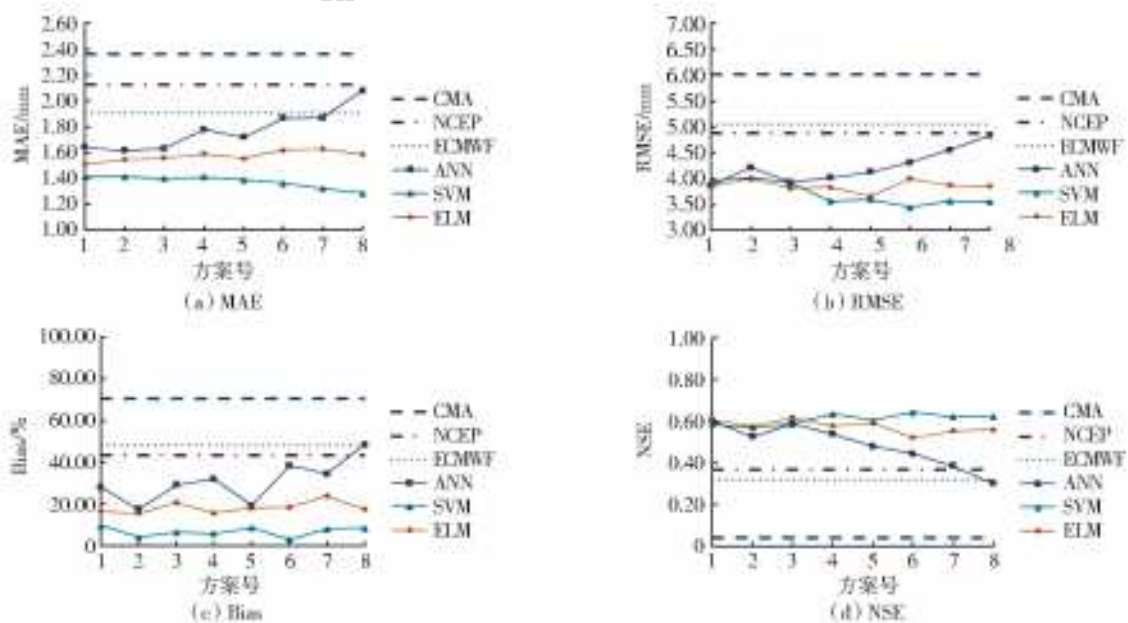


图 2 单一模式和集成预报模型验证期未来第 2 天降雨预报精度对比

Fig. 2 Comparison of the forecast accuracy of rainfall for the next second day between single ensemble forecast center and three integrated forecasting schemes in verification period

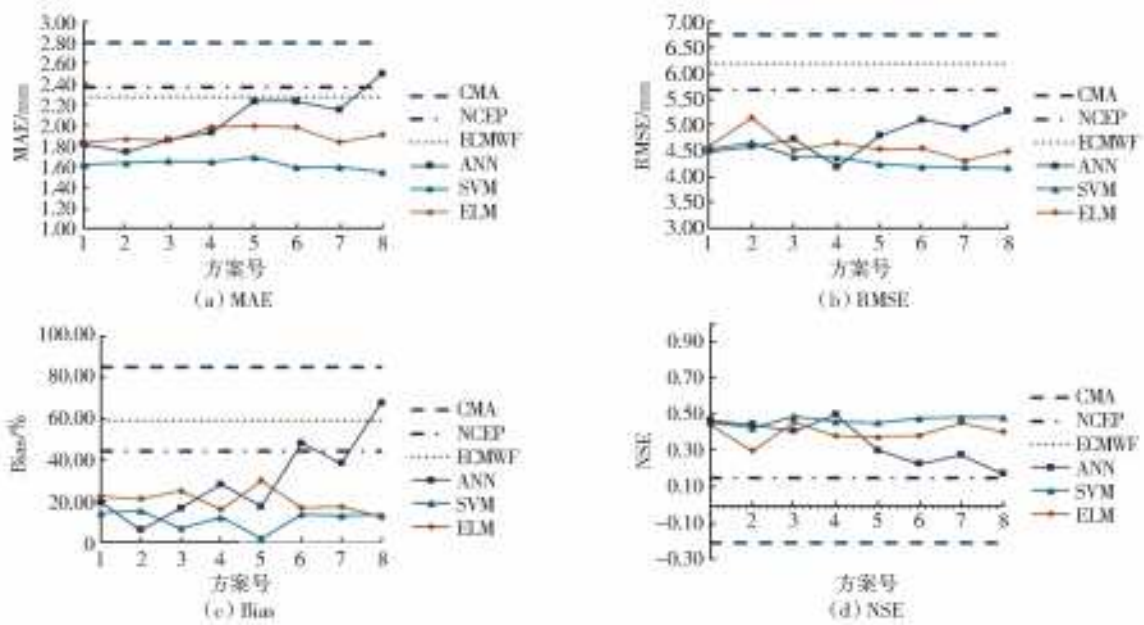


图 3 单一模式和集成预报模型验证期未来第 3 天降雨预报精度对比

Fig. 3 Comparison of the forecast accuracy of rainfall for the next third day between single ensemble forecast center and three integrated forecasting schemes in verification periods

2.4.1 预报精度分析

图 1 至图 3 给出了验证期内各单一模式和集成预报模型未来 1~3 d 预报降雨与实测降雨之间的绝对平均误差 MAE、均方根误差 RMSE、相对误差 Bias 以及纳什系数 NSE。可以看出,对于未来 1~3 d,在不同的输入因子组合下,SVM 和 ELM 模型的预报效果明显优于各单一预报中心,预报效果比较稳定,且通过比较各指标可以发现,SVM 的预报效果优于 ELM。对于 ANN 模型,预报未来第 1 天降雨时,其预报效果表现出了一定的不稳定性,从图 1 中可以看出,当方案号为 2,3,4,5 时,其预报效果各项指标均优于单一模式,其它方案号时则出现了评价指标劣于单一模式的情况;预报未来第 2、3 天降雨时(图 2、图 3),当输入因子数量较少时,其各项指标均优于各单一模式,当输入因子数量增加到一定数量时,部分指标将出现劣于单一模式的情况(如图 2 中方案 8 以及图 3 中方案 6),且各项指标波动较大。这是因为 ANN 模型对输入因子的影响较为敏感,当优质输入因子被引入时,预报效果将得到改善,当劣质输入因子被引入时,预报效果将受到破坏,随着 n 的增加,优质因子和劣质因子被随机引入模型,预报效果呈现出一定的波动性。因此,输入因子选择合适时,ANN 模型仍然可以提高降雨预报精度。

2.4.2 预报准确率分析

上述结果表明,集成预报与单一模式相比,其预报精度更高。在实际应用中,调度人员同时关心各

量级降雨的预报准确率。因此,本文进一步对集成预报方法预报结果进行准确率分析,表 2 给出了验证期内各单一模式未来 1~3 d 各级预报降雨准确率,表 3 至 5 给出了验证期内集成预报模型各方案的预报准确率。从表 2 至表 5 中可以看出,对于未来 1~3 d 内 II、III、IV 预报降雨,SVM 和 ELM 模型相对于各单一模式,其预报准确率提高较为明显,例如就未来第 1 天的 III 级降雨而言,单一模式最大准确率为 41%(见表 2),而 SVM 模型各方案中,准确率最小值为 57%,最大为 65%,平均值为 61%;ELM 模型各方案中,准确率最小值为 51%,最大值为 58%,平均值为 54%,均大于 41%(见表 3)。就未来第 3 天的 IV 级降雨而言,单一模式的最大准确率为 34%(见表 2),而 SVM 和 ELM 模型的最小准确率分别为 50%和 40%(见表 5),均大于单一模式的最大准确率。对于未来 1~3 d 内 I 级预报降雨,SVM 和 ELM 模型与各单一模式的预报准确率水平相当,甚至出现下降的情况,例如针对未来第 1 天的 I 级降雨,单一模式的预报准确率最大值和最小值分别为 92%和 87%,而 SVM 模型各方案的最大值和最小值分别为 89%和 88%;ELM 模型的最大值和最小值分别为 91%和 82%,这是因为对于各单一模式而言,I 级降雨预报准确率已经很高,不同预报中心预报信息的集成反而使训练数据更加复杂,模型模拟难度增加,预报效果很难继续改善。

对于 ANN 模型,其预报准确率多数情况下优于单一模式,但也呈现出一定的不稳定性。随着方案

表 2 各单一模式未来 1~3 d 各级降雨预报准确率

Tab. 2 Forecast accuracy of rainfall at all levels over the next 1-3 days for each single model 单位: %

降雨等级	未来第 1 天			未来第 2 天			未来第 3 天		
	ECMWF	CMA	NCEP	ECMWF	CMA	NCEP	ECMWF	CMA	NCEP
I	91	87	92	89	84	91	88	83	88
II	56	52	55	56	49	53	55	47	52
III	41	31	37	40	29	30	29	30	25
IV	44	39	31	33	44	50	26	34	33

表 3 集成预报模型在验证期未来第 1 天的各级降雨预报准确率

Tab. 3 The accuracy of rainfall forecasting at all levels for the next day in the verification period using integrated forecasting models 单位: %

模型	降雨等级	1	2	3	4	5	6	7	8	最大	最小	平均
ANN	I	90	87	90	92	89	92	90	90	92	87	90
	II	60	64	60	57	61	56	61	63	64	56	60
	III	46	48	51	47	54	47	58	43	58	43	49
	IV	41	59	50	30	56	29	42	16	59	16	40
SVM	I	88	88	88	89	88	89	88	88	89	88	88
	II	66	66	66	68	61	60	61	71	71	60	65
	III	59	59	60	57	58	63	63	65	65	57	61
	IV	60	60	60	60	50	44	50	60	60	44	56
ELM	I	91	90	87	88	85	86	82	87	91	82	87
	II	60	61	66	71	66	67	61	69	71	60	65
	III	51	53	52	51	55	58	58	55	58	51	54
	IV	44	55	56	63	50	50	60	67	67	44	56

注:表中加粗部分为出现集成预报模型准确率小于单一模式准确率的情况。

表 4 集成预报模型在验证期未来第 2 天的各级降雨预报准确率

Tab. 4 The accuracy of rainfall forecasting at all levels for the next second day in the verification period using integrated forecasting models 单位: %

模型	降雨等级	1	2	3	4	5	6	7	8	最大	最小	平均
ANN	I	91	87	87	89	88	83	86	90	91	83	88
	II	52	62	56	53	57	61	64	56	64	52	57
	III	50	46	47	49	44	54	41	43	54	41	47
	IV	60	60	50	58	50	29	17	12	60	12	42
SVM	I	91	91	88	91	89	91	87	85	91	85	89
	II	56	57	61	57	58	56	65	60	65	56	59
	III	52	51	50	60	60	68	70	59	70	50	59
	IV	71	63	60	71	50	71	50	57	71	50	62
ELM	I	90	90	88	87	89	87	86	86	90	86	88
	II	56	59	61	58	56	59	63	64	64	56	60
	III	48	48	47	55	57	56	60	68	68	47	55
	IV	63	71	60	67	56	63	50	50	71	50	60

注:表中加粗部分为出现集成预报模型准确率小于单一模式准确率的情况。

号的增加, II、III、IV 级降雨预报准确率出现了小于单一模式的少数情况(表 3 至 5 中加粗部分), 例如表 3 中方案号为 4 时, ANN 模型 IV 级预报降雨的准确率为 30%, 小于 ECMWF、CMA 和 NCEP 的 44%、39% 和 31%。这是由于输入因子对模型预报效果的影响, 一方面, 随着方案号的增加, 输入因子越来越多, 输入信息的冗余使得 ANN 模型的模拟和预报精度有所降低, 另一方面, 劣质输入因子的引入也会降低模型预报效果, 与 2.4.1 中分析原因一致。因此

在选择 ANN 模型进行集成预报时, 需要注意输入因子的选择。

为了更加直观地比较集成模型与单一模式的预报准确率水平, 表 6 给出了单一模式和集成模型在不同预见期、不同降雨等级条件下的预报准确率最大值与平均值。从表 6 中可以直观地看出, 对于 I 级预报降雨, 集成预报模型与单一模式预报准确率差别不大, 预见期为 1 d 时, 单一模式和集成模型准确率最大值均为 92%, 平均值分别为 90% 和 88%, 预见期为

表 5 集成预报模型在验证期未来第 3 天的各级降雨预报准确率

Tab. 5 The accuracy of rainfall forecasting at all levels for the next third day in the verification period using integrated forecasting models 单位: %

模型	降雨等级	1	2	3	4	5	6	7	8	最大	最小	平均值
ANN	I	90	87	81	87	63	84	81	79	90	63	81
	II	55	60	66	58	39	55	63	41	66	39	55
	III	43	36	33	47	50	38	42	44	50	33	42
	IV	50	60	63	75	50	29	29	24	75	24	47
SVM	I	88	89	90	90	89	90	89	90	90	88	89
	II	59	58	57	55	54	54	57	59	59	54	57
	III	36	39	33	53	57	62	64	70	70	33	52
	IV	50	50	50	50	50	50	50	50	50	50	50
ELM	I	90	91	87	85	83	85	85	84	91	83	86
	II	57	58	59	55	56	56	60	56	60	55	57
	III	37	39	35	42	46	40	60	55	60	35	44
	IV	40	50	57	67	60	50	40	50	67	40	52

注:表中加粗部分为出现集成预报模型准确率小于单一模式准确率的情况。

表 6 集成预报模型各方案准确率平均值和最大值与单一预报模式比较

Tab. 6 Comparison of the average and maximum value of accuracy between the integrated forecasting models and the single forecasting models 单位: %

预见期	降雨等级	最大值		平均值	
		单一模式	集成模型	单一模式	集成模型
1	I	92	92	90	88
	II	56	71	54	63
	III	41	65	36	55
	IV	44	67	38	51
2	I	91	91	88	88
	II	56	65	53	59
	III	40	70	33	54
	IV	50	71	42	55
3	I	88	91	86	85
	II	55	66	51	56
	III	30	70	28	46
	IV	34	75	31	50

注:单一模式的准确率最大值和平均值指的是在预见期和降雨等级相同的条件下,ECMWF、CMA 和 NCEP 三个单一模式中,预报准确率的最大值和平均值;集成模型的准确率最大值和平均值指的是在预见期和降雨等级相同的条件下,ELM、ANN 和 SVM 模型各方案中,预报准确率的最大值和平均值。

2 d 时,单一模式和集成模型准确率最大值均为 91%,平均值均为 88%。对于 II、III、IV 级预报降雨,集成预报方法预报准确率要显著高于各单一预报中心,就 IV 级降雨而言,预见期为 1 d 时,单一模式和集成模型准确率最大值分别为 44%和 67%,平均值分别为 38%和 51%,预见期为 2 d 时,单一模式和集成模型准确率最大值分别为 50%和 71%,平均值分别为 42%和 55%。由此可见,基于机器学习模型的多模式集成预报技术可行,能够提高桓仁水库流域短期预报降雨准确率。

3 结 论

为了研究基于机器学习模型的多模式降雨集成预报技术的可利用性,本文以 ECMWF、CMA 以及 NCEP 三个单一集合预报中心 2011 年 1 月 1 日至 2017 年 12 月 31 日每日及前期发布的预报降雨为输入因子,驱动 ELM、ANN 以及 SVM 模型,对桓仁流域未来 1~3 d 降雨进行集成预报,并从预报精度和预报准确率角度分析了预报效果,初步得出以下结论。

(1)对于不同的输入因子组合方案,SVM 和 ELM 模型预报效果均优于单一模式,即各项精度评价指标和各级降雨预报准确率均优于单一模式;ANN 模型则在多数情况下(输入因子选择合适时)优于单一模式,说明基于机器学习模型的多模式降雨集成预报技术在桓仁流域可行,能够减小未来 1~3 d 内降雨预报误差。

(2)就预报效果而言,不同模型预报效果不尽相同,其中 SVM 模型在三种模型中效果最优、ELM 次之,ANN 模型最差。因此,模型的选择会影响集成预报方法的效果,推荐使用 SVM 模型。

(3)各集成预报模型的预报效果对于输入因子具有一定的依赖性,即相同的模型在不同的输入因子组合条件下会有不同的预报效果,特别是 ANN 模型,对输入因子最为敏感。因此在进行多模式降雨集成预报时,输入因子需要科学的选择。

参考文献(References):

- [1] COLLISCHONN W, MORELLI TUCCI C E, CLARKE R T, et al. Medium-range reservoir inflow predictions

- based on quantitative precipitation forecasts [J]. Journal of Hydrology. 2007, 344(1-2): 112-122. DOI: 10.1016/j.jhydrol.2007.06.025.
- [2] CLOKE H L, PAPPENBERGER F. Evaluating forecasts of extreme events for hydrological applications; an approach for screening unfamiliar performance measures [J]. Meteorological Applications. 2008, 15(1): 181-197. DOI: 10.1002/met.58.
- [3] LEE C, HO H, LEE K T, et al. Assessment of sewer flooding model based on ensemble quantitative precipitation forecast [J]. Journal of Hydrology. 2013(506): 101-113. DOI: 10.1016/j.jhydrol.2012.09.053.
- [4] 彭勇,徐炜,王萍,等. 耦合 TIGGE 降水集合预报的洪水预报[J]. 天津大学学报(自然科学与工程技术版), 2015, 48(2): 177-184. (PENG Y, XU W, WANG P, et al. Flood forecasting coupled with TIGGE ensemble precipitation forecasts [J]. Journal of Tianjin University, 2015, 48(2): 177-184. (in Chinese)) DOI: 10.11784/tdxbz201308048.
- [5] 钟伟,朱从飞,张卫国,等. 基于 SCE-UA 算法的多模式降雨集成预报技术研究[J]. 宁波大学学报(理工版), 2018, 31(6): 27-31. (ZHONG W, ZHU C F, ZHANG W G, et al. Multi-model integrated precipitation forecast based on SCE-UA algorithm [J]. Journal of Ningbo University (Natural Science & Engineering Edition), 2018, 31(6): 27-31. (in Chinese)) DOI: CNKI: SUN: NBDZ. 0. 2018-06-005.
- [6] 赵成武. 短时降水的超级集合预报算法研究[D]. 长沙: 国防科学技术大学研究生院, 2014. (ZHAO C W. Study of the algorithm for super-ensemble in short-range precipitation forecast [D]. Changsha: Graduate school of National University of Defense Technology, 2014. (in Chinese)) DOI: CNKI: CDMD: 2. 1016. 921643.
- [7] CARTWRIGHT T J, KRISHNAMURTI T N. Warm season mesoscale superensemble precipitation forecasts in the Southeastern United States [J]. Weather and Forecasting. 2007, 22(4): 873-886. DOI: 10.1175/waf1023.1.
- [8] WU J, LU G, WU Z. Flood forecasts based on multi-model ensemble precipitation forecasting using a coupled atmospheric-hydrological modeling system [J]. Natural Hazards. 2014, 74(2): 325-340. DOI: 10.1007/s11069-014-1204-6.
- [9] LIM Y, JO S, LEE J, et al. Multimodel ensemble forecasting of rainfall over East Asia: regularized regression approach [J]. International Journal of Climatology. 2014, 34(14): 3720-3731. DOI: 10.1002/joc.3938.
- [10] 王海霞,智协飞. 基于 TIGGE 资料的降尺度及多模式集成降雨预报研究[A]. 创新驱动发展 提高气象灾害防御能力——S3 第三届气象服务发展论坛[C]. 公众、专业气象预报服务技术与应用, 2013: 679-690. (WANG H X, ZHI X F. Research on downscale and multi-model integrated rainfall forecast based on TIGGE data [A]. Innovation-driven development Improving the ability of meteorological disaster prevention—the third meteorological service development forum of S3-public [C]. professional meteorological forecast service technology and application. 2013: 679-690. (in Chinese)).
- [11] 黄威,牛若芸. 基于集合预报和支持向量机的中期强降雨集成预报试验 [J]. 气象, 2017, 43(9): 1110-1116. (HUANG W, NIU R Y. The medium-term multi-model integration forecast experimentation for heavy rain based on support vector machine [J]. Meteorological Monthly. 2017, 43(9): 1110-1116. (in Chinese)) DOI: 10.7519/j.issn.1000-0526.2017.09.008.
- [12] YASEEN Z M, JAAFAR O, DEO R C, et al. Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq [J]. Journal of Hydrology. 2016(542): 603-614. DOI: 10.1016/j.jhydrol.2016.09.035.
- [13] ZHANG X L, PENG Y, ZHANG C, et al. Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting Some experiment evidences [J]. Journal of Hydrology. 2015, 530: 137-152. DOI: 10.1016/j.jhydrol.2015.09.047.
- [14] 冯鑫伟,黄领梅,沈冰. 多元线性回归与 BP 神经网络模型在次洪量预测中的对比研究 [J]. 水资源与水工程学报, 2017, 28(3): 123-126. (FENG X W, HUANG L M, SHEN B. Comparative study on multivariate linear regression and BP neural network model in the prediction of flood volume [J]. Journal of Water Resources and Water Engineering, 2017, 28(3): 123-126. (in Chinese)) DOI: 10.11705/j.issn.1672-643X.2017.03.23.
- [15] 聂敏,刘志辉,刘洋,等. 基于 PCA 和 BP 神经网络的径流预测 [J]. 中国沙漠, 2016, 36(4): 1144-1152. (NIE M, LIU Z H, LIU Y, et al. Runoff forecast based on principal component analysis and BP neural network [J]. Journal of Desert Research, 2016, 36(4): 1144-1152. (in Chinese)) DOI: 10.7522/j.issn.1000-694X.2015.00039.

- [16] 明波,刘冀,吕翠美,等.基于小波分析的径流分级组合预报模型[J].人民长江,2012,43(17):61-64. (MING B,LIU J,LYU C M, et al. Classification and combination prediction model of runoff based on wavelet analysis[J]. Yangtze River, 2012, 43(17): 61-64. (in Chinese)) DOI: 10. 3969/j. issn. 1001-4179. 2012. 17. 016.
- [17] RAMAN H, CHANDRAMOULI V. Deriving a general operating policy for reservoirs using neural network [J]. Journal of Water Resources Planning and Management, 1996, 122(5): 342-347.
- [18] CORINNA C, VAPNIK V. Support-vector networks [J]. Machine Learning, 1995, 20(3): 273-297.
- [19] 赵红标,吴义斌.基于支持向量机的中长期入库径流预报[J].红水河,2009,28(5):55-59. (ZHAO H B, WU Y B. Long-term runoff forecast based on support vector machine[J]. Hongshui River, 2009, 28(5): 55-59. (in Chinese)) DOI: 10. 3969/j. issn. 2095-008X. 2009. 03. 001.
- [20] VAPNIK V N. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995. DOI: 10. 1007/978-1-4757-2440-0.
- [21] 庄楚强,何春雄.应用数理统计基础[M].广州:华南理工大学出版社,2013. (ZHUANG C Q, HE C X. The basis of applied mathematical statistics [M]. Guangzhou: South China University of Technology Press, 2013. (in Chinese))
- [22] HUANG G B, WANG D H, LAN Y. Extreme learning machines: A survey [J]. International Journal of Machine Learning and Cybernetics, 2011, 2(2): 107-122. DOI: 10. 1007/s13042-011-0019-y.
- [23] HUANG G B, ZHU Q, SIEW C. Extreme learning machine: Theory and applications [J]. Neurocomputing, 2006, 70(1-3): 489-501. DOI: 10. 1016/j. neucom. 2005. 12. 126.
- [24] HUANG G B, ZHOU H, DING X, et al. Extreme learning machine for regression and multiclass classification [J]. IEEE Trans Syst Man Cybern B Cybern, 2012, 42(2): 513-529. DOI: 10. 1109/tsmcb. 2011. 2168604.
- [25] 颜宏文,盛成功.基于层次聚类和极限学习机的母线短期负荷预测[J].计算机应用,2018,38(8):2437-2441. (YAN H W, SHENG C G. Short-term bus load forecasting based on hierarchical clustering method and extreme learning machine [J]. Journal of Computer Applications, 2018, 38(8): 2437-2441. (in Chinese)) DOI: 10. 11772/j. issn. 1001-9081. 2018010017.
- [26] 王小川,史峰,郁磊. MATLAB 神经网络 43 个案例分析 [M]. 北京:北京航空航天大学出版社,2013. (WANG X C, SHI F, YU L. 43 cases analysis of matlab neural network [M]. Beijing: Beihang University Press, 2013. (in Chinese))
- [27] 崔慧慧,智协飞.基于 TIGGE 资料的地面气温延伸期多模式集成预报[J].大气科学学报,2013,36(2):165-173. (CUI H H, ZHI X F. Multi-model ensemble forecasts of surface air temperature in the extended range using the TIGGE dataset [J]. Journal of Nanjing Institute of Meteorology, 2013, 36(2): 165-173. (in Chinese)) DOI: 10. 3969/j. issn. 1674-7097. 2013. 02. 005.

• 科普知识 •

谈黄河的防洪问题

黄河中游暴雨频繁、强度大、历时短,洪水具有洪峰高、历时短、陡涨陡落的特点。目前黄河流域建成水利工程总库容达 900 亿 m^3 ,有效库容超过 450 亿 m^3 ,形成了比较完整的防洪工程体系,但是洪水问题不可能彻底解决。1843 年黄河花园口站洪峰流量达到 33 000 m^3/s ,目前花园口过流能力最大为 22 000 m^3/s ,黄河下游最窄处过流能力只有 11 000 m^3/s 。所以未来下游城市群发展、滩区治理等都得考虑未来遭遇超标准洪水的可能性,做好防洪预案,建立洪水预报预警系统,保护好人民群众的生命财产安全。

——节选自韩宇平《专家学者谈黄河·多维调控是解决黄河水问题的重要方略》,南水北调与水利科技微信公众平台,2019 年 11 月。