

文章编号: 1007-2284(2020)10-0021-08

# 不同降水降尺度方法 在天山西部区域的适用性评估

喻雪晴 穆振侠 周育琳

(新疆农业大学水利与土木工程学院, 乌鲁木齐 830052)

**摘要:** 为能更好地提高外源数据的可靠性与适用性, 丰富基础资料匮乏地区可用的数据源, 以天山西部区域的伊宁、尼勒克和昭苏站作为典型代表站, 基于实测降水数据和 NCEP/NCAR 再分析数据, 采用人工神经网络 ANN、极限学习机 ELM、长短期记忆神经网络 LSTM、支持向量机 SVM、组合简单平均 CSAM 和组合支持向量机 CSVM 方法建立统计降尺度模型, 评估不同降尺度方法在降水降尺度方面的适用性。研究结果表明: ①由于降水影响因素的多重性、实测站点降水的差异性, 不同降尺度方法的优劣性存在一定的差异, 且率定期与验证期最优降尺度方法可能出现不对应的情况; ②实测站点降水量越大, 降尺度效果越好; 降水量年内变化越不均匀, 降尺度效果越差; ③由于实测站点冬季降雪事件的发生影响了降尺度的效果, 总体降雪量越多, 降尺度效果越差; ④极端降水事件频发, 影响到降尺度的效果, 降低了降尺度结果的可靠性与精度。总体上, 研究区 ELM、LSTM 和 CSVM 模型具有一定的优越性, 其中 CSVM 组合模型在季节、年降水方面模拟效果最好, ELM 模型在月降水分布方面模拟效果最好。

**关键词:** 降尺度; ANN; ELM; LSTM; 组合模型

**中图分类号:** P339; TV273

**文献标识码:** A

## An Assessment of Different Precipitation Downscaling Methods in the Western Tianshan Mountains

YU Xue-qing, MU Zhen-xia, ZHOU Yu-lin

(College of Water Conservancy and Civil Engineering, Xinjiang Agricultural University, Urumqi 830052, China)

**Abstract:** In order to better improve the reliability and applicability of external data and enrich the data sources available in areas where basic data are scarce. The Yining, Nilek and Zhaosu stations in the western Tianshan region are taken as typical representative stations. Then, the statistical downscaling model is established based on measured precipitation data and NCEP/NCAR reanalysis data by using a series of methods such as artificial neural network (ANN), extreme learning machine (ELM), long and short term memory neural network (LSTM), support vector machine (SVM), combined simple average (CSAM) and combined support vector machine (CSVM). Finally, the applicability of different downscaling methods of precipitation downscaling are evaluated. The results show that: ① Due to the multiplicity of precipitation influencing factors and the difference of precipitation at measured stations, there are some differences in the advantages and disadvantages of different downscaling methods. What's more, the optimal downscaling method of the calibration period may not correspond to the optimal downscaling method of the validation period. ② The greater the precipitation of the measured site, the better the downscaling effect. The more uneven the precipitation during the year, the worse the downscaling effect. ③ Because the occurrence of snowfall events in winter affects the effect of downscaling, the more the overall snowfall, the worse the effect of downscaling. ④ The frequent occurrence of extreme precipitation events affects the effect of downscaling and reduces the reliability and accuracy of downscaling results. Generally speaking, the ELM, LSTM and CSVM models in the study area have certain advantages, among which the CSVM combination model has the best simulation effect on seasonal and annual precipitation, and the ELM model has the best simulation effect on the monthly precipitation distribution.

**Key words:** downscaling; ANN; ELM; LSTM; combination model

收稿日期: 2020-01-15

基金项目: 新疆维吾尔自治区自然科学基金项目(2018D01A16)

**作者简介:** 喻雪晴(1995-), 女, 硕士研究生, 主要从事水文水资源方向的研究。E-mail: 1207218841@qq.com。

**通讯作者:** 穆振侠(1980-), 男, 教授, 主要从事水文水资源方向的研究。E-mail: muzhenxia@126.com。

气温、降水等水文气象要素对水资源合理开发利用、洪旱灾害预警、社会经济可持续发展、生态安全的维护等方面具有至关重要的作用。但目前部分区域因自然条件、技术手段的限制,在获取基础水文气象数据方面存在一定问题,致使无法更好地为区域水文预报、径流规律研究等服务,为此迫切需要拓展基础数据源。伴随着遥感卫星数据和地理信息系统的快速发展,为拓展数据源方面提供了更多的可能性<sup>[1]</sup>。但这些外源数据受到设计框架和处理速度的制约,其空间分辨率过于粗糙无法满足精度要求,尤其是在反映区域尺度水文气象要素方面具有较大的局限性,不能直接用于小尺度区域的水文水资源研究中<sup>[2]</sup>。为解决该问题,急切需要寻找可行的方法,基于已有的基础数据对拓展的外源数据进行降尺度,以提高外源数据的空间分辨率、可靠性与适用性。对于地形地貌极其复杂的新疆地区,水文气象观测站点稀少,大部分观测站点布设于平原区及出山口位置,且分布密度严重不足。很多中高山带地区的基础水文气象资料严重匮乏甚至没有资料<sup>[3-7]</sup>,而这些中高山带地区是山区河流的发源地及径流的形成区,决定着所在区域水资源量的多少<sup>[8]</sup>。所以本研究对于丰富资料匮乏地区可用的数据源及提高降水精度方面具有重要的意义。

近二十年来,人们对降尺度方法及其应用进行了广泛的研究,许多技术和方法被提出和使用,可以分为统计降尺度方法和动力降尺度方法。其中统计降尺度方法灵活,很大程度节约了计算机运行时间,且能更有效的解决大尺度气候要素和小尺度气候要素空间尺度不匹配问题,相较于动力降尺度法应用更加广泛。它主要是建立预报因子和预报量之间的统计关系,并用独立的观测资料检验这种关系的合理性<sup>[2,9-12]</sup>,常用的统计降尺度方法有:人工神经网络法<sup>[13-15]</sup>、极限学习机法<sup>[15,16]</sup>、支持向量机法<sup>[17-21]</sup>、长短期记忆神经网络法<sup>[22]</sup>等。大部分研究表明统计降尺度方法对气温的模拟效果好、精度高,但对降水的模拟效果不太理想,无法满足精度要求<sup>[8,23,24]</sup>。因此,本文以天山西部区域作为研究区,选取有不同特点的气象站点实测降水数据集以及美国国家环境预报中心/国家大气研究中心再分析数据集,在单一统计降尺度方法适用性及优劣分析的基础上,如人工神经网络 ANN( Artificial Neural Network)、极限学习机 ELM( Extreme Learning Machine)、长短期记忆神经网络 LSTM( Long and Short Term Memory Neural Network)和支持向量机 SVM( Support Vector Machine),采用组合降尺度方法分析在研究区的适用性及可行性。具体组合降尺度方法有组合简单平均法 CSAM( Combining Simple Average Method)和组合支持向量机法 CSVM( Combining Support Vector Machine)。简单平均法 CSAM 是对 ANN、ELM 和 LSTM 3 种统计降尺度模型的降尺度结果进行简单算术平均,得到一个组合的降水模拟结果。组合支持向量机 CSVM 是将 ANN、ELM 和 LSTM 3 种统计降尺度模型的降尺度结果作为输入数据带入支持向量机模型中,输出一组新的降水模拟结果。

## 1 数据和方法

### 1.1 数据来源

#### 1.1.1 实测数据

为了更好地研究不同降尺度方法在天山西部区域的适用性

评估,结合研究区已建基础站点的特点,有针对性地选择参与降尺度分析的基础站点。结合研究区降水时空变化规律特点,在不同站点降水规律分析的基础上,选取冬季降水量偏多的伊宁站、年降水量最大的昭苏站及靠近山区的尼勒克站作为典型代表站,参与降尺度分析。收集伊宁、尼勒克、昭苏气象站点 1981–2010 年的实测日降水量序列资料,数据来源于中国气象科学数据共享服务网。

#### 1.1.2 NCEP/NCAR 再分析数据

NCEP/NCAR 再分析数据是由美国国家环境预报中心和大气中心联合制作的,是全球再分析数据中发展最早、时间尺度最长的数据,可提供自 1948 年至今的预报产品。它是在气候模式输出数据的基础上,采用最先进的同化系统(三维变分同化技术)对地面观测数据和卫星遥感等数据进行同化处理,得到的全球气象资料数据库,具有很强的可靠性,目前被广泛应用于统计降尺度等相关研究<sup>[2,6,17-20]</sup>。本研究收集 1981–2010 年空间分辨率为  $2.5^\circ \times 2.5^\circ$ 、时间分辨率为日的 NCEP/NCAR 再分析资料,并根据以往研究选取对降水有显著性影响的因子作为备选的预报因子变量,如表 1 所示。

表 1 备选预报因子变量

Tab.1 Alternative forecasting factors utilized in this study

序号	预报因子	描述	序号	预报因子	描述
1	mslp	平均海平面气压	8	uwnd	近地表纬向风速
2	rhum	近地表相对湿度	9	v500	500 hPa 径向风速
3	t500	500 hPa 温度	10	v850	850 hPa 径向风速
4	t850	850 hPa 温度	11	vwnd	近地表径向风速
5	temp	近地表温度	12	w500	500 hPa 垂直速度
6	u500	500 hPa 纬向风速	13	z500	500 hPa 位势高度
7	u850	850 hPa 纬向风速	14	z850	850 hPa 位势高度

## 1.2 研究方法

### 1.2.1 预报因子选择

选择合适的预报因子是降尺度过程中最重要的步骤,可以避免因预报因子过多造成维数过高、过于拟合等问题。因此,本文借助 NCEP/NCAR 再分析资料的 14 个备选预报因子,采用逐步回归分析法<sup>[25,26]</sup>对预报因子进行筛选。14 个备选预报因子和降水量分别作为自变量和因变量,先将全部自变量逐个引入回归模型,每一个新自变量引入后进行  $F$  检验,并对已经选入的自变量逐个进行  $t$  检验,当原始引入的变量因为添加变量后不再显著,就删除后引入的变量,重复这个过程可以保证回归模型中只包含显著变量,此时回归模型为最优方程模型,这不仅能有效减少预报因子的数量,降低统计降尺度模型的复杂程度和计算量。

### 1.2.2 统计降尺度方法

为了较好地降低外源数据的空间分辨率以补充研究区基础气象数据,本文借助实测降水数据和逐步回归分析法筛选得到的预报因子,分别建立人工神经网络 ANN<sup>[13-15]</sup>、极限学习机 ELM<sup>[15,16]</sup>、长短期记忆神经网络 LSTM<sup>[22]</sup>和支持向量机 SVM<sup>[17-21]</sup>四种单一统计降尺度模型,以及组合简单平均 CSAM 和组合支持向量机 CSVM 的组合统计降尺度模型。以 1981–

2000 年率定期、2001–2010 年为验证期开展分析研究。

### 1.2.3 模型的评价指标

为了更好地验证不同降水降尺度方法在天山西部区域的适用性、更好的评价不同降尺度方法优劣性,本研究分别从 3 个方面入手进行检验不同降尺度方法的模拟结果,一是检验降水实测值与模拟值之间相关关系的密切程度,选用确定系数( $R^2$ )作为统计指标;二是检验降水实测值与模拟值之间的偏差程度,选用均方根误差( $RMSE$ )作为统计指标;三是检验各模型的可信程度,选用纳什效率系数( $NS$ )作为统计指标。计算公式如下:

$$R^2 = \left[ \frac{\sum_{i=1}^N (O_i - \bar{O}_i)(P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^N (O_i - \bar{O}_i)^2 \sum_{i=1}^N (P_i - \bar{P}_i)^2}} \right]^2 \quad (1)$$

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

$$NS = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2} \quad (3)$$

式中: $O_i$ 为实测降水值,mm; $P_i$ 为模拟降水值,mm; $N$ 为样本总数; $\bar{O}_i$ 为实测降水值平均,mm; $\bar{P}_i$ 为模拟降水值平均,mm。

其中 $R^2$ 值越接近于 1,表明降水量模拟值波动越接近实测值波动,两者相关关系的密切程度越高。 $RMSE$ 值越接近于 0,表明降水模拟值与实测值之间的偏差程度越小。 $NS$ 值越接近于 1,表明模型的可信程度越高质量越好; $NS$ 值为 0,表明实测值与模拟值大致接近,但存在显著的误差; $NS$ 值远远小于 0,表明模型是不可信的。

## 2 结果与分析

为更好地指导地形复杂、降水规律多变及影响因素众多区域水资源的合理开发利用、防灾减灾工作的开展及气候变化的应对等与降水相关方面的工作,本研究从冬季降水量偏多、年降水量最大及靠近山区角度出发选取典型降尺度站点,并借助 NCEP/NCAR 再分析数据,在合理预报因子筛选的基础上,分别借助 4 种单一统计降尺度模型(ANN、ELM、LSTM 和 SVM)与两种组合降尺度模型(CSAM 和 CSVM),探讨了不同降尺度方法在研究区的适用性。为避免由于过多预报因子增加计算维度,复杂或增加计算量,产生过度拟合及更多不确定的问题,本研究采用逐步回归分析法,从显著性水平方面分别对典型站点优选 6 个预报因子,进行不同降尺度方法的适用性研究。为更好地克服降尺度方法优劣反映的片面性,在研究过程中对样本数据划定率定期(1981–2000 年)与验证期(2001–2010 年),从降尺度结果与实测数据的相关性、偏差程度、各模型的可信程度角度出发,引入 $R^2$ 、 $RMSE$ 、 $NS$ 指标及降水累积分布函数(CDF),分别从月、季、年的尺度探讨了单一与组合降尺度方法对选用典型站点过程与总量反映的优劣性,并从不同的角度推荐可行的降尺度方法。

### 2.1 预报因子的筛选结果

借助 SPSS 软件将样本数据作为被解释变量,待备选的 14 个预报因子作为解释变量,对每一个所考虑的解释变量进行逐步回归分析。每引入一个解释变量后都要进行 $F$ 检验,并对已经选入的解释变量逐个进行 $t$ 检验,当原来引入的解释变量由于后面解释变量的引入变得不再显著时,则将其删除,最终从备选的 14 个预报因子中筛选出显著性最高且满足 $t$ 检验的 6 个因子作为本研究进行降尺度分析的预报因子。伊宁站, rhum、t500、t850、v500、vwnd 和 z500 因子与降水最显著,经 $F$ 检验, $F=31.078$ , $P=0<0.001$ ;尼勒克站,mslp、rhum、t500、temp、v850、z850 因子与降水最显著,经 $F$ 检验, $F=59.192$ , $P=0<0.001$ ;昭苏站,t500、t850、u850、uwnd、w500、z850 因子与降水最显著,经 $F$ 检验, $F=148.216$ , $P=0<0.001$ ;3 个站点筛选的 6 个因子经 $t$ 检验发现 $p$ 的值都在标准值 0.1 以下。相对来说,对于不同站点选择的大尺度预报因子有所不同,但大都选择温度、径向风速和位势高度变量作为最终预报因子,表明温度、径向风速和位势高度是影响降水量必不可少的关键因素。

### 2.2 不同模型降尺度结果

#### 2.2.1 不同方法优劣性分析

基于 ANN、ELM、LSTM、SVM、CSAM 和 CSVM 方法,选用 1981–2000 年的伊宁、尼勒克和昭苏站点实测降水量数据与逐步回归分析筛选得到的预报因子率定各统计降尺度模型,选用 2001–2010 年的数据验证各统计降尺度模型。分别计算确定系数( $R^2$ )、均方根误差( $RMSE$ )、纳什效率系数( $NS$ )来评价各统计降尺度模型率定和验证过程中对降水的模拟效果,具体如表 2 所示。总体可以看出:①年降水量越大(多年平均降水量

表 2 率定期和验证期统计指标结果

Tab.2 Statistical values of calibration and validation

站点	模型	率定期			验证期		
		$R^2$	$RMSE/mm$	$NS$	$R^2$	$RMSE/mm$	$NS$
伊宁	ANN	0.310	12.712	0.308	0.348	14.248	0.332
	ELM	0.381	12.041	0.379	0.426	13.349	0.414
	LSTM	0.375	12.093	0.374	0.394	13.702	0.382
	SVM	0.366	12.168	0.366	0.402	13.728	0.380
	CSAM	0.392	11.919	0.392	0.400	13.685	0.384
	CSVM	0.418	11.683	0.416	0.430	13.521	0.398
尼勒克	ANN	0.528	16.471	0.522	0.488	18.350	0.479
	ELM	0.556	15.886	0.555	0.566	17.007	0.553
	LSTM	0.547	16.032	0.547	0.534	17.637	0.519
	SVM	0.529	16.393	0.526	0.512	17.909	0.504
	CSAM	0.559	15.856	0.557	0.534	17.531	0.525
	CSVM	0.567	15.699	0.565	0.534	17.500	0.527
昭苏	ANN	0.778	18.256	0.765	0.762	18.927	0.744
	ELM	0.775	18.216	0.766	0.770	18.386	0.759
	LSTM	0.772	18.197	0.766	0.789	17.710	0.776
	SVM	0.775	18.057	0.770	0.780	17.832	0.773
	CSAM	0.780	17.926	0.773	0.781	18.184	0.764
	CSVM	0.781	17.810	0.776	0.785	18.147	0.765

伊宁站为 298.81 mm、尼勒克站为 406.88 mm、昭苏站为 507.07 mm) 不同降尺度方法降尺度效果越好; ②由于不同站点降水年际变化较大、不均匀程度较高(变差系数  $C_v$  在 1981–2010、1981–2000、2001–2010 3 个阶段伊宁站分别为 0.72、0.68、0.79, 尼勒克站 0.73、0.75、0.71, 昭苏站为 0.93、0.97、0.88), 不同降尺度方法在率定期与验证期存在最佳降尺度方法不一致的情况; ③由于选用站点冬季降雪量较大(伊宁站月降雪达为 114.80 mm, 发生在 2004 年 11 月, 占年降水量的 23.1%; 尼勒克站为 74 mm 发生在 2004 年 11 月, 占年降水量的 17.8%; 昭苏站 37.3 mm 发生在 1994 年 11 月, 占年降水量的 7.8%), 影响到不同选用站点降尺度的效果, 总体规律为冬季或月份降雪量越小, 降尺度效果越好, 反之越差; ④由于 2000 年后研究区易发生极端降水事件, 降水在年内变化极不均匀, 影响到了降尺度的效果, 增加了降尺度的难度, 其中伊宁站 2000 年后反应降水不均匀变化的变差系数  $C_v$  呈明显增大趋势、尼勒克站变差系数  $C_v$  呈弱的减小趋势、昭苏站呈明显的减小趋势, 也说明了不同站点降尺度效果的优劣的原因。一般来说, 变差系数  $C_v$  减小、降水年内变化越均匀, 降水降尺度效果呈变好的趋势, 但不同降尺度方法在不同站点体现的效果存在一定的差异。

基于筛选的预报因子, 借助不同的降尺度方法选取典型站点进行不同降尺度方法优劣的对比。率定期, 由于选取站点年降水的差异, 总体效果以降水量较大的昭苏站不同方法降尺度效果最好, 其次为降水次之的尼勒克站, 伊宁站效果较差。伊宁站不同降尺度方法的确定性系数  $R^2$  在 0.310~0.418 之间、均方根误差  $RMSE$  在 11.683~12.712 mm 之间、纳什效率系数  $NS$  在 0.308~0.416 之间; 从单降尺度模型角度考虑以 ELM 最优、LSTM 次之, 后为 SVM; 从组合降尺度的方法而言以 CSVM 最优、CSAM 次之; 从综合的角度而言, 以组合降尺度的方法 CSVM 在伊宁站率定期降尺度效果最好。尼勒克站不同降尺度方法的确定性系数  $R^2$  在 0.528~0.567 之间、均方根误差  $RMSE$  在 15.699~16.471 mm 之间、纳什效率系数  $NS$  在 0.522~0.565 之间; 从单降尺度模型角度考虑以 ELM 最优、LSTM 次之, 后为 SVM; 从组合降尺度的方法而言以 CSVM 最优、CSAM 次之; 从综合的角度而言, 以组合降尺度的方法 CSVM 在尼勒克站率定期降尺度效果最好。昭苏站不同降尺度方法的确定性系数  $R^2$  在 0.772~0.781 之间、均方根误差  $RMSE$  在 17.810~18.256 mm 之间、纳什效率系数  $NS$  在 0.765~0.776 之间; 从单降尺度模型角度考虑以 SVM 最优、LSTM 次之, 后为 ELM; 从组合降尺度的方法而言以 CSVM 最优、CSAM 次之; 从综合的角度而言, 以组合降尺度的方法 CSVM 在昭苏站率定期降尺度效果最好。

由于在验证期 2001–2010 年时间范围内, 研究区极端降水事件的出现、降水年际不均匀情况的变化, 造成了率定期与验证期不同降尺度方法效果的差异性与不对应性, 给降尺度方法的研究与成果的应用带来了一定的挑战, 但总体还是年降水量越少, 降尺度效果越差。伊宁站验证期不同降尺度方法的确定性系数  $R^2$  在 0.348~0.430 之间、均方根误差  $RMSE$  在 13.349~14.248 mm 之间、纳什效率系数  $NS$  在 0.332~0.414 之间; 从单降尺度模型角度考虑以 ELM 最优、LSTM 次之, 后为 SVM; 从组合降尺度的方法而言以 CSVM 最优、CSAM 次之; 从综合的角度

而言, 以单降尺度模型 ELM 在伊宁站验证期降尺度效果最好。尼勒克站验证期不同降尺度方法的确定性系数  $R^2$  在 0.488~0.566 之间、均方根误差  $RMSE$  在 17.007~18.350 mm 之间、纳什效率系数  $NS$  在 0.479~0.553 之间; 从单降尺度模型角度考虑以 ELM 最优、LSTM 次之, 后为 SVM; 从组合降尺度的方法而言以 CSVM 最优、CSAM 次之; 从综合的角度而言, 以单降尺度模型 ELM 在尼勒克站验证期降尺度效果最好。昭苏站验证期不同降尺度方法的确定性系数  $R^2$  在 0.762~0.789 之间、均方根误差  $RMSE$  在 17.710~18.927 mm 之间、纳什效率系数  $NS$  在 0.744~0.776 之间; 从单降尺度模型角度考虑以 LSTM 最优、SVM 次之, 后为 ELM; 从组合降尺度的方法而言以 CSVM 最优、CSAM 次之; 从综合的角度而言, 以单降尺度模型 LSTM 在昭苏站验证期降尺度效果最好。

综合率定期与验证期的结果可知, 由于选用站点的典型性、降水变化的复杂性, 总体而言, 所选的降尺度方法基本上能够模拟研究区的降水情况, 这与不同研究区对于降尺度方法方面的研究规律与效果基本一致。由于不同的站点影响降水的因素各不相同, 正常情况下各站点的不同降尺度方法优劣性是存在差异的, 但是考虑到实用的方便性和实际的可操作性, 从区域综合的角度来看, 研究区 3 个代表站点率定期以组合降尺度的方法 CSVM 降尺度效果最好, 而验证期以单降尺度模型 ELM 与 LSTM 效果最好。

不同降尺度方法得到的模拟降水与实测降水之间的误差如表 3 所示。可以看出, 各站点大都低估了 4、5、6、7、10 和 11 月的降水, 低估范围在 0.15~13.45 mm 之间, 而大都高估了 1、2、3、8、9 和 12 月的降水, 高估范围介于 -0.06~ -3.93 mm 之间, 这是由于研究区域降水量主要集中在 4~7 月份, 各模型难以模拟到极值降水, 出现了低估强降水的情况, 而对于 12 月和 1~3 月, 表现出高估弱降水的情况; 从单站点来看, ANN 单模型对伊宁站月降水量模拟效果最好, ELM 模型对尼勒克站和昭苏站月降水量模拟效果最好; 从区域综合的角度来看, 研究区 3 个代表站点使用 ELM 模型得到的模拟降水与实测降水之间的误差较小, 模拟效果好。

不同降尺度方法得到的模拟年降水与实测年降水之间的误差如图 1 所示。从单站点来看, 各模型模拟的伊宁站年降水误差主要集中在 -10~15 mm 之间, 其中单降尺度模型以 ELM 最优、LSTM 次之, 组合降尺度模型以 CSVM 最优; 尼勒克站年降水误差主要集中在 -15~20 mm 之间, 其中单降尺度模型以 ELM 最优、LSTM 次之, 组合降尺度模型以 CSVM 最优; 昭苏站年降水误差主要集中在 -18~22 mm 之间, 其中单降尺度模型以 LSTM 最优, ELM 次之, 组合降尺度模型以 CSVM 最优。从区域综合的角度来看, 研究区 3 个代表站点使用 CSVM 组合模型得到的误差小、模拟效果较好。具体哪一种降尺度的方法更适用于研究区, 还需从降水的月分布、季节、年变化过程、相对误差大小等方面进行综合比选, 以更好开展研究降水降尺度方面的研究与成果的应用。

### 2.2.2 月降水分布变化规律分析

为了更直观、更全面地评估研究区的最佳降尺度方法, 针对率定期与验证期综合表现最好的单降尺度模型 ELM 与 LSTM、

表 3 不同降尺度方法的误差对比  
Tab.3 Error comparison of different downscaling methods

站点	方法	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月	平均
伊宁	ANN	-3.23	-2.85	-0.62	1.49	1.89	5.66	5.21	-5.93	-6.54	3.94	6.54	-1.60	0.33
	ELM	-2.79	-1.89	-0.69	-0.28	0.69	5.92	6.47	-1.62	-4.25	4.08	6.49	-1.42	0.89
	LSTM	-2.09	-1.38	0.67	2.02	2.39	8.09	8.04	-2.13	-3.93	6.26	7.43	-0.43	2.08
	SVM	-1.43	-0.85	1.03	3.70	4.61	9.26	9.47	-0.12	-3.53	6.45	8.13	0.52	3.10
	CSAM	-2.70	-2.04	-0.21	1.08	1.66	6.56	6.58	-3.22	-4.91	4.76	6.82	-1.15	1.10
	CSVM	-1.53	-0.46	0.75	0.96	2.03	8.20	8.65	0.21	-2.35	5.81	7.71	-0.40	2.46
尼勒克	ANN	1.70	-1.57	-0.06	3.41	7.56	10.24	8.13	-1.51	-0.83	3.77	3.22	0.97	2.92
	ELM	0.29	-0.26	-1.68	1.31	4.20	3.86	4.86	-2.04	-3.48	3.90	3.93	-1.01	1.16
	LSTM	-1.78	-2.12	0.17	1.79	3.50	6.13	9.18	0.87	-0.79	4.77	4.43	1.60	2.31
	SVM	0.06	-0.92	0.60	1.69	4.62	8.29	10.58	0.57	-0.64	4.74	4.69	1.82	3.01
	CSAM	0.07	-1.31	-0.52	2.17	5.09	6.74	7.39	-0.89	-1.70	4.15	3.86	0.52	2.13
	CSVM	1.02	1.04	1.22	0.69	3.03	4.38	6.07	-0.77	-0.27	5.46	5.24	2.12	2.44
昭苏	ANN	-1.03	-0.14	0.45	2.82	11.18	2.71	8.66	7.04	0.07	0.46	2.00	-1.36	2.74
	ELM	-0.89	0.46	-1.46	0.31	10.31	0.15	8.69	7.47	-3.01	0.32	2.32	-0.76	1.99
	LSTM	-3.58	-3.93	-2.37	1.75	10.30	3.86	13.45	7.19	1.39	3.47	-0.73	-3.50	2.27
	SVM	-1.05	-0.54	-2.30	-0.62	9.90	2.14	12.01	7.73	-1.80	2.20	2.12	-1.08	2.39
	CSAM	-1.83	-1.20	-1.13	1.63	10.60	2.24	10.27	7.23	-0.52	1.42	1.20	-1.87	2.34
	CSVM	-2.90	-2.36	0.28	4.16	8.86	0.31	8.44	7.51	2.65	5.42	1.81	-2.28	2.66

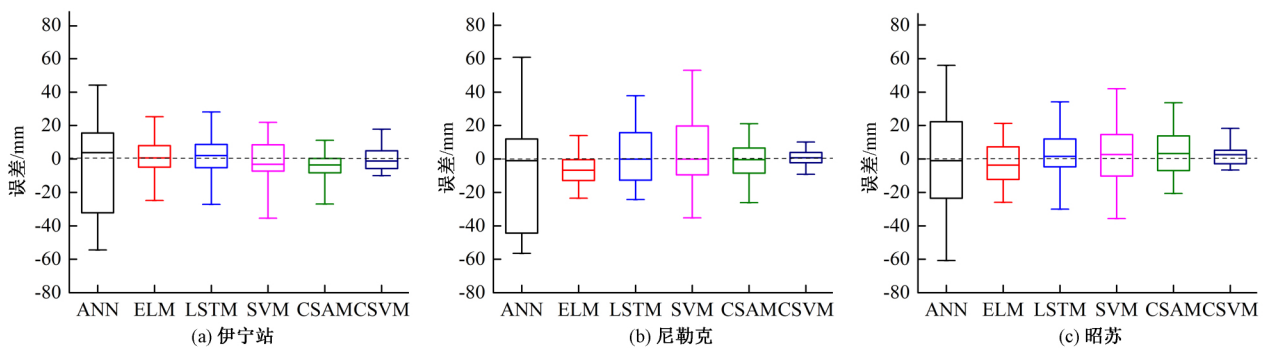


图 1 不同降尺度方法的年误差图

Fig.1 Error of precipitation

组合降尺度模型 CSVM 采用月降水累积分布函数( CDF) 分析特定站点不同降尺度方法对于月降水变化规律反映的正确性。图 2 为伊宁站 1981-2010 年 ELM、LSTM 和 CSVM 模型月降水累积分布函数( CDF) 的对比情况。如 1 月份实测降水阈值小于 45 mm 时,降水累积概率已超过 95%, 这表明 1 月份发生 45 mm 以下的降水概率较高, 而出现 45 mm 或以上的降水概率较低; ELM、LSTM 和 CSVM 模型模拟的降水阈值均小于 30 mm 时,降水累积概率超过 95%, 表明降尺度结果难以模拟到极大降水, 且模拟降水值偏小, 类似情况也发生在其余月份。又如 7 月份实测降水阈值小于 10 mm 时,降水累积概率已超过 25%, 而 ELM、LSTM 和 CSVM 模型均没有模拟到小于 10 mm 的降水量, 表明降尺度结果也较难捕捉到极小降水, 类似情况同样较为明显的体现在 4、8、9 和 12 月。尽管 3 种统计降尺度方法均能较好地模拟出各月降水量的累积分布规律, 但是 3 种统计降

尺度方法的累积分布函数图之间没有特别明显的差异性, 在尼勒克站和昭苏站也同样得到类似的结果, 考虑到篇幅原因, 仅列出了伊宁站不同降尺度模型 CDF 图。基于累积分布函数, 采用 Kolmogorov-Smirnov 非参数拟合优度检验( K-S 检验) 来对比分析 3 种统计降尺度方法模拟月降水累积分布和实测月降水累积分布之间的相似性, 以通过 95% 置信水平(  $P=0.05$ ) 作为评估标准( 表 4)。可以看出, 伊宁站不同降尺度模型主要在 4 月、7 和 8 月未通过 95% 置信水平( 其中 ELM 模型 7 月未通过 95% 置信水平、LSTM 模型 7 和 8 月两个月未通过检验、CSVM 组合模型 4、7 和 8 月未通过检验), 表明特定的月份不同降尺度方法降尺度结果难以再现这些月份的降水分布。尼勒克站, ELM 和 CSVM 组合模型全年 12 个月降尺度结果均通过 95% 置信水平检验, 说明 ELM 和 CSVM 组合模型很好地重现了尼勒克月降水量分布情况; 而 LSTM 模型在 1、2 和 12 月未通过

$P=0.05$  的置信检验表明 LSTM 模型的降尺度结果难以再现这些月份的降水分布。昭苏站,ELM 模型全年 12 个月均通过  $P=0.05$  的置信检验,说明 ELM 模型很好地重现了昭苏月降水分布情况;LSTM 模型在 3 月、CSVM 组合模型在 1、2 和 3 月未通过  $P=0.05$  的置信检验,无法再现这些月份的降水分布情况。从单站点来看,ELM 模型对伊宁月降水量分布的模拟效果最好,LSTM 模型次之,后为 CSVM 组合模型;ELM 模型和 CSVM

组合模型对尼勒克月降水量分布的模拟效果最好,LSTM 模型较差;ELM 单模型对昭苏月降水量分布的模拟效果最好,LSTM 模型次之,后为 CSVM 组合模型。从区域综合的角度来看,ELM 模型比其他两种方法能较好地再现研究区 3 个代表站点月降水量的分布情况,但对于 LSTM 模型和 CSVM 组合模型重现的月降水量分布,CSVM 组合模型的模拟效果略优于 LSTM 模型。

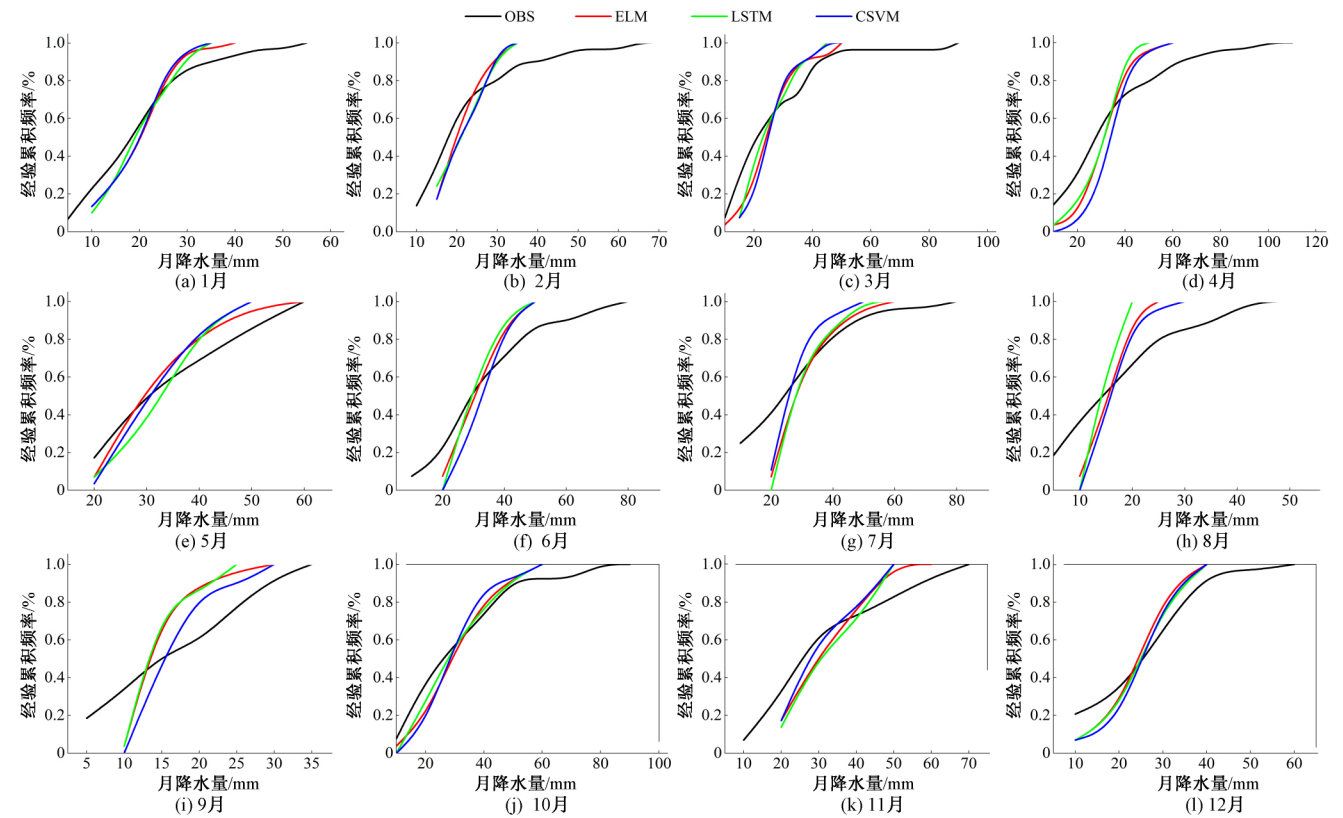


图2 伊宁站月降水累积分布函数对比图

Fig.2 Comparison of cumulative distribution function of monthly precipitation in Yining Station

表4 K-S 检验结果( $p$  值)

Tab.4 K-S test results ( $p$  values)

月份	伊宁			尼勒克			昭苏		
	ELM	LSTM	CSVM	ELM	LSTM	CSVM	ELM	LSTM	CSVM
1	0.537	0.760	0.760	0.153	0.036	0.153	0.153	0.078	0
2	0.321	0.321	0.321	0.300	0.030	0.088	0.983	0.096	0.017
3	0.279	0.279	0.078	0.200	0.200	0.342	0.098	0.004	0.009
4	0.300	0.300	0.042	0.098	0.098	0.098	0.514	0.927	0.741
5	0.184	0.321	0.098	0.200	0.109	0.055	0.537	0.342	0.537
6	0.466	0.279	0.078	0.490	0.490	0.169	0.200	0.109	0.109
7	0.019	0.019	0.019	0.109	0.200	0.055	0.109	0.055	0.200
8	0.078	0.002	0.006	0.200	0.109	0.200	0.109	0.055	0.109
9	0.078	0.078	0.078	0.514	0.184	0.514	0.109	0.342	0.342
10	0.258	0.440	0.138	0.169	0.169	0.169	0.537	0.537	0.537
11	0.321	0.514	0.514	0.200	0.055	0.537	0.741	0.184	0.098
12	0.321	0.514	0.184	0.279	0.016	0.153	0.892	0.258	0.068



### 2.2.3 季节降水变化规律分析

为更好的探究优选的不同降尺度模型的优劣性,对 ELM、LSTM 和 CSVM 组合模型降尺度结果在季节上与实际情况相符性采用模拟与实测季节降水量间平均绝对误差(MAE)指标进行分析,如图3所示。从单站点来看,3种降尺度方法对伊宁站冬季降水量的模拟效果最佳,其次是夏季和秋季,后为春季,其中以 CSVM 组合模型模拟效果最优,ELM 模型次之,后为 LSTM 模型;对尼勒克站冬季降水量的模拟效果最佳,秋季次之,其次是春季,后为夏季,其中以 CSVM 组合模型模拟效果最优,ELM 模型次之,后为 LSTM 模型;对昭苏站冬季降水量的模拟效果

最佳,秋季次之,其次是春季,后为夏季,其中以 CSVM 组合模型模拟效果最优,ELM 模型次之,后为 LSTM 模型。从区域综合的角度来看,CSVM 组合模型对研究区3个代表站点季节降水量模拟效果最佳,平均绝对误差值大都集中在4~18 mm之间,具体如图3(d)所示,对于伊宁(YN)、尼勒克(NLK)和昭苏(ZS)站点使用 CSVM 组合模型模拟的冬季降水量效果最好,秋季次之,其次是春季,夏季模拟效果较差。这是因为相较于其余季节,冬季降水量小,故模拟得到的平均绝对误差值小,也说明对降水量越小的季节,降尺度模拟效果越好。

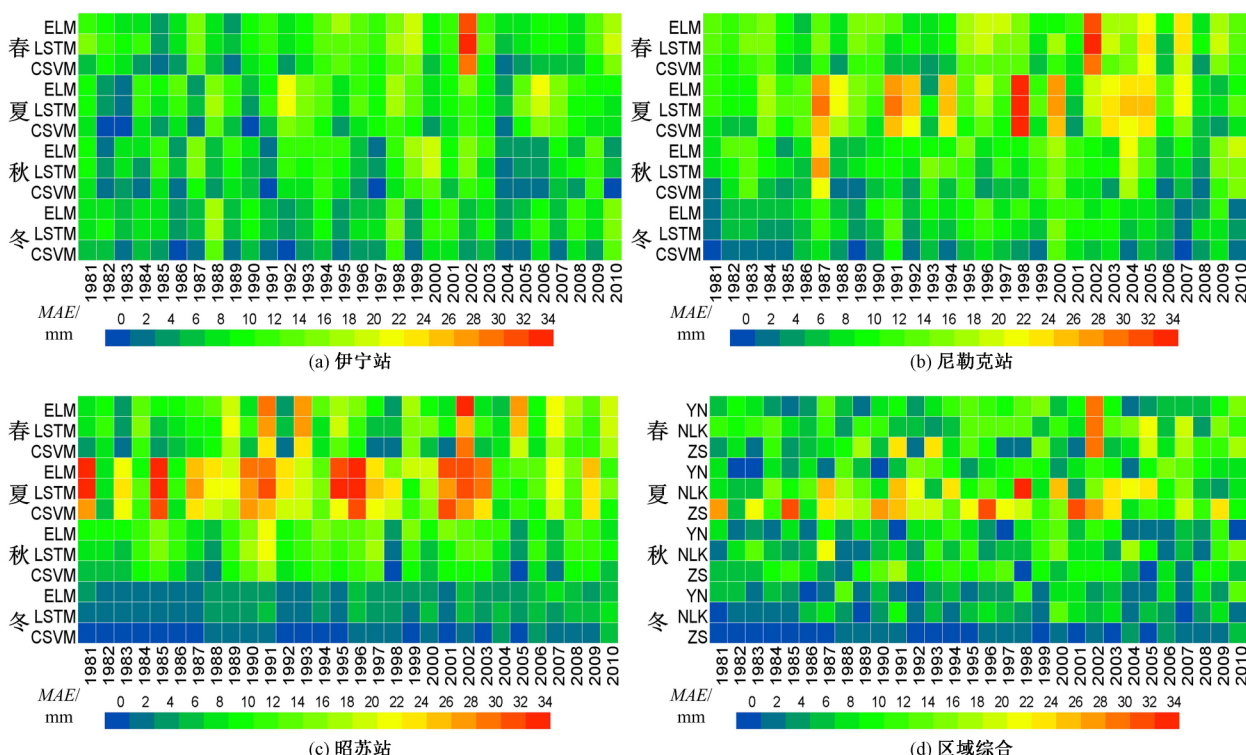


图3 季节降水量平均绝对误差热图

Fig.3 Heatmap of seasonal precipitation mean absolute error

### 2.2.4 年降水变化规律分析

通过比较 ELM 模型、LSTM 模型和 CSVM 组合模型的模拟降水量与实测降水量间平均相对误差,对降水量的年变化进行规律分析。如图4所示,伊宁站3种统计降尺度方法的平均相对误差均集中在-2%~8%,其中 CSVM 组合模型模拟效果最好,ELM 单模型次之,后为 LSTM 单模型。尼勒克站3种统计

降尺度方法的平均相对误差均集中在-3%~4%,其中 CSVM 组合模型模拟效果最好,LSTM 单模型次之,后为 ELM 单模型。昭苏站3种统计降尺度方法的平均相对误差均集中在-2%~2%,其中 CSVM 组合模型模拟效果最好,ELM 单模型次之,后为 LSTM 单模型。总体来说,3种统计降尺度方法模拟的平均相对误差,昭苏站最好,尼勒克站次之,后为伊宁站;CSVM 组合模型的模拟效果最佳,其次是 ELM 单模型,LSTM 单模型较差。

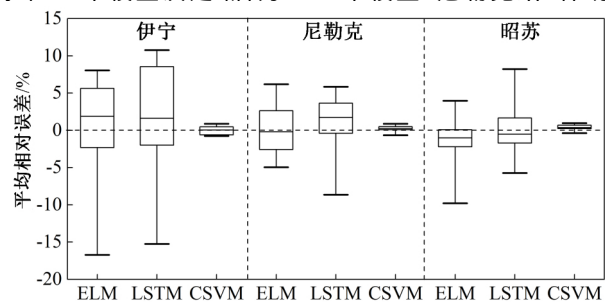


图4 年降水量平均相对误差

Fig.4 Annual average relative error of precipitation

## 3 结论

本文采用逐步回归分析法筛选预报因子,基于伊宁、尼勒克和昭苏站点1981-2010年实测降水数据和筛选的预报因子分别建立人工神经网络 ANN、极限学习机 ELM、长短期记忆神经网络 LSTM 和支持向量机 SVM 的单一统计降尺度模型,以及组合简单平均 CSAM 和组合支持向量机 CSVM 的组合统计降尺度模型,对研究区典型站点降水进行降尺度分析,使用3种统计指标对各统计降尺度模型的模拟能力进行评估,并从月、

季节和年降水变化等方面进行综合比选。得到以下主要结论。

(1) 从不同方法优劣性来看,利用逐步回归分析筛选得到的预报因子与实测降水数据所建立的 6 种统计降尺度模型均适用于模拟研究区降水;各模型均对年降水量最大的昭苏站模拟效果最优,尼勒克站次之,后为伊宁站;CSVM 组合模型在率定期的降尺度效果最好,单降尺度模型 ELM 与 LSTM 在验证期的降尺度效果最好。

(2) 从月降水变化来看,ELM 单模型、LSTM 单模型和 CSVM 组合模型均能大致的模拟出各月降水量的累积分布规律,但降尺度结果较难再现极端降水量;3 种统计降尺度方法模拟的大多数月份的月降水量分布均通过了 Kolmogorov-Smirnov 非参数拟合优度检验,置信水平  $P$  值都达到 0.05 以上;单降尺度模型 ELM 对再现各站点月降水量的分布情况表现最好,CSVM 组合模型模拟效果略优于 LSTM 单模型。

(3) 从季节降水变化来看,ELM 单模型、LSTM 单模型和 CSVM 组合模型对冬季降水量的模拟效果最优,其次是春季和秋季,对夏季降水量的模拟效果相对较差;CSVM 组合模型的平均绝对误差相对较小,对季节降水量的模拟效果最优,ELM 单模型次之,后为 LSTM 单模型。

(4) 从年降水变化来看,3 种统计降尺度方法均表现出对昭苏站的降水模拟效果最好,尼勒克站次之,后为伊宁站;CSVM 组合模型对年降水的模拟效果最佳,ELM 单模型次之,后为 LSTM 单模型。

由于选取降尺度站点的特殊性及其降水年内、年际变化的差异性,引起了降尺度效果的优劣性差异、率定与验证期最佳方法的不一致性以及降尺度方法的不唯一性。为更好地服务与生产实践,还需从降水规律本身的特点、影响因素、形成规律等方面进行数据规律的挖掘,并探究适应的降尺度方法。

通过对选用的 6 种统计降尺度模型分析发现,ELM、LSTM 和 CSVM 模型相比较而言有一定的优越性,能较好地适用于本研究区的降水降尺度方面的研究与成果的应用。CSVM 组合模型在季节、年降水方面模拟效果最好,ELM 模型在月降水分布方面模拟效果最好。但模拟降水量与实测降水量之间仍存在一定的偏差,该偏差主要是因为降尺度模型难以捕捉到极值降水,即便极值降水出现的情况较少,但模拟极值降水是制约模型模拟精确度的关键。在今后的研究中,如果能充分考虑极值降水,并贴切的考虑研究区的实际情况建立降尺度模型,模型的模拟精确度可能会进一步提高,以便更好地补充研究区基础水文气象资料、水文风险预报和径流规律模拟研究等方面的工作。 □

#### 参考文献:

- [1] 汤瑞,穆振侠,周育琳,等.基于 SRM 模型的天山西部山区融雪径流研究[J].灌溉排水学报,2018,37(11):106-112.
- [2] 褚健婷,夏军,许崇育. SDSM 模型在海河流域统计降尺度研究中的适用性分析[J].资源科学,2008,30(12):1825-1832.
- [3] 夏依拉·居尔艾提,杨涛.资料匮乏地区径流降尺度模型构建及预测[J].中国农村水利水电,2016(1):12-15.
- [4] 于宴民,穆振侠. CFSR 数据在高寒山区径流模拟中的适用性[J].灌溉排水学报,2015,34(11):93-97.
- [5] 高瑞,穆振侠.天山西部山区 VIC 模型的应用[J].南水北调与水利科技,2017,15(4):44-48.
- [6] 高瑞,穆振侠,彭亮,等.水量平衡与能量平衡模式下的 VIC 模型在喀什河流域的应用研究[J].灌溉排水学报,2018,37(1):106-112.
- [7] 汤瑞,穆振侠,内孜买买提·阿不来提,等.基于不同方法在天山西部山区雪线的确定及其变化规律分析[J].南水北调与水利科技,2018,16(2):108-114.
- [8] 穆振侠,姜卉芳.基于统计降尺度方法的高寒山区未来气候变化预估[J].干旱区研究,2015,32(2):290-296.
- [9] 郭靖,郭生练,陈华,等.丹江口水库未来径流变化趋势预测研究[J].南水北调与水利科技,2008,8(4):78-82.
- [10] 向田恬,陈华,郭家力,等.气候变化对嘉陵江流域降水变化影响分析[J].南水北调与水利科技,2010,8(1):75-91.
- [11] 张明月,彭定志,胡林涓.统计降尺度方法研究进展综述[J].南水北调与水利科技,2013,11(3):118-122.
- [12] 张静文,郭家力,刘佳,等.鄱阳湖流域入湖径流时空变化预测[J].南水北调与水利科技,2016,14(5):41-48.
- [13] YANG C, WANG N, WANG S, et al. Performance comparison of three predictor selection methods for statistical downscaling of daily precipitation[J]. Theor Appl Climatol, 2018(131):43-54.
- [14] Osman Y Z, Abdellatif M E. Improving accuracy of downscaling rainfall by combining predictions of different statistical downscale models[J]. Water Science, 2016(30):61-75.
- [15] Alizamir M, Azhdary Moghadam M, Hashemi Monfared A, et al. Statistical downscaling of global climate model outputs to monthly precipitation via extreme learning machine: A case study[J]. Environmental Progress & Sustainable Energy, 2018(37):1853-1862.
- [16] 孔德萌,李维德,吴金冉.基于协整理论的极限学习机模型在降水预测中的应用[J].水电能源科学,2017,35(9):1-3.
- [17] 刘向培,王汉杰,何明元.应用统计降尺度方法预估江淮流域未来降水[J].水科学进展,2012,23(1):29-37.
- [18] Liu J, Yuan D, Zhang L, et al. Comparison of three statistical downscaling methods and ensemble downscaling method based on bayesian model averaging in upper Hanjiang river basin, China[J]. Advances in Meteorology, 2016:1-12.
- [19] Sa'adi Z, Shahid S, Chung E S, et al. Projection of spatial and temporal changes of rainfall in Sarawak of Borneo Island using statistical downscaling of CMIP5 models[J]. Atmospheric Research, 2017(197):446-460.
- [20] Pour S H, Shahid S, Chung E S, et al. Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of Bangladesh[J]. Atmospheric Research, 2018(213):149-162.
- [21] Pour S H, Shahid S, Chung E S. A Hybrid Model for Statistical Downscaling of Daily Rainfall[J]. Procedia Engineering, 2016(154):1424-1430.
- [22] 冯钧,潘飞.一种 LSTM-BP 多模型组合水文预报方法[J].计算机与现代化,2018,275(7):86-92.
- [23] 刘绿柳,刘兆飞,徐宗学.21 世纪黄河流域中游地区气候变化趋势分析[J].气候变化研究进展,2008,4(3):167-172.
- [24] 翟文亮,李朋俊,林凯荣,等.基于 SDSM-SWAT 的气候变化下东江流域径流预测模拟[J].人民珠江,2016,37(4):1-6.
- [25] 章杰,白红英,袁博,等.秦岭地区气温变化统计降尺度研究[J].干旱区研究,2013,30(2):322-328.
- [26] 许燕,王世杰,白晓永,等.基于 SDSM 的珠江中上游气候模拟及未来情景预估[J].中国岩溶,2018,37(2):72-81.