

# 基于混合模型的中长期降水量预测

李 栋<sup>1</sup> 薛惠锋<sup>1,2</sup>

(西北工业大学自动化学院 西安 710072)<sup>1</sup> (中国航天系统科学与工程研究院 北京 100048)<sup>2</sup>

**摘 要** 针对中长期降水量预测精度较低的问题,提出了由改进集合经验模态分解方法、最小二乘法、核极限学习机和改进的果蝇优化算法构成的混合模型来对区域年度降水量序列进行预测。首先,通过改进集合经验模态分解方法将非平稳降水量时间序列分解为多个分解项。然后,根据不同分解项的特性分别采用最小二乘法和核极限学习机对其进行预测。由于核极限学习机均存在一定的参数敏感特性,因此提出使用改进的果蝇优化算法来对核极限学习机的相关参数搜索寻优,以提高其预测精度。最后,将各分解项的预测结果叠加,从而形成最终预测结果。以广东省 7 个地市 1951—2015 年的年度降水量为例,对所提方法进行了验证,结果表明:相比于自回归移动平均模型和核极限学习机模型,混合模型预测具有更高的预测精度。

**关键词** 预测,混合模型,改进集合经验模态分解方法,最小二乘法,核极限学习机,改进果蝇优化算法

**中图分类号** TP391.9 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.09.045

## Forecasting of Medium and Long Term Precipitation Based on Hybrid Model

LI Dong<sup>1</sup> XUE Hui-feng<sup>1,2</sup>

(School of Automation, Northwestern Polytechnical University, Xi'an 710072, China)<sup>1</sup>

(China Aerospace Academy of Systems Science and Engineering, Beijing 100048, China)<sup>2</sup>

**Abstract** Accurate estimation of precipitation is an important issue in water resources engineering, management and planning. In order to improve the accuracy of medium and long term precipitation forecasting, a hybrid forecasting model based on modified ensemble empirical mode decomposition, least squares method, kernel extreme learning machine and modified fruit fly optimization algorithm was presented. By using modified ensemble empirical mode decomposition, non-stationary precipitation time series is decomposed into many terms. Then the decomposed terms are predicted by the least square method or the kernel extreme learning machine according to its characteristics. Because the kernel extreme learning machine has some characteristics of parameter sensibility, the modified fruit fly optimization algorithm is used to search the optimal parameters to improve the forecasting accuracy. Finally, forecast results of each decomposed term are added together to obtain the final forecasting results. The method was tested by using annual precipitation data from seven cities in China's Guangdong province between 1951 and 2015. Results show that compared with the auto-regressive moving average model and kernel extreme learning machine model, the mixed model has higher prediction accuracy.

**Keywords** Prediction, Hybrid model, Modified ensemble empirical mode decomposition, Least squares method, Kernel extreme learning machine, Modified fruit fly optimization algorithm

## 1 引言

水与人类的生产、生活息息相关,是地区和国家发展的不可或缺的基础性资源。随着我国经济的飞速发展和人口基数的不断增长,生活用水与工业用水的需求量在不断增大,水资源危机凸显,加强对水资源的管理已经迫在眉睫。水资源管理涉及水资源的计划、开发与分配,而这些都与降水量预测有着直接或者间接的关系<sup>[1]</sup>。因此,正确识别降水的演变模式并有效预测降水量,对于地区、国家开展水资源的优化配置具有重要的意义。

长期以来,国内外学者对降水量预测进行了多角度探索,并提出了多种预测方法,这些方法总体而言可以分为三大类:成因预测方法、统计预测方法和机器学习方法。1)成因预测方法主要通过降水天气过程的研究识别其影响因素(如温度、湿度、压力、风力等),继而构建成因模型<sup>[2-5]</sup>。近年来,随着人类活动对各地气候的影响越来越大,大气降水的影响因子的识别难度增加,通过成因来预测中长期降水量的预测精度难以取得满意的结果。2)统计预测方法根据降水历史数据,利用非线性时间序列分析<sup>[6]</sup>、ARIMA<sup>[7]</sup>等方法来构建统计模型。由于统计预测方法大多忽视了气候系统非线性、非

平稳性的特点,仅仅依靠原始数据序列来预测降水量未来的动态变化使得预测精度普遍不高。3)机器学习方法利用神经网络<sup>[8]</sup>等机器学习算法构建预测模型。然而,中长期降水量的随机性和复杂性会大大降低神经网络的收敛速度和预测精度,甚至会造成不收敛<sup>[8]</sup>,从而导致预测结果不稳定且预测精度较差。总之,传统的成因预测方法、统计预测方法以及单一的机器学习预测方法均存在一定的局限性,难以对复杂的非线性降水序列进行准确的预测。

近年来,一些新的数据分析技术相继出现并在气象领域得到应用。例如,由 Huang 等提出的经验模态分解方法(Empirical Mode Decomposition, EMD)能将一个非平稳序列分解为若干平稳序列,从而提高了其可预测性,目前该方法已经在降水量预测<sup>[6,9-10]</sup>、气温分析<sup>[11]</sup>中得到了应用。随着 EMD 研究的逐步深入,研究人员发现 EMD 在自适应分解过程中会不可避免地产生模态混合现象,这种模态混合不仅使得各 IMF 无法准确展现,甚至导致个别 IMF 丧失其物理意义。为了尽可能地降低模态混合的影响, Huang 等提出在分解过程中引入白噪声,并通过集合平均的方法降低白噪声的影响,这个方法被称为集合经验模态分解<sup>[12]</sup>(Ensemble Empirical Mode Decomposition, EEMD)方法。EEMD 方法能够有效降低模态混合对分解的影响,但由于在引入白噪声时,噪声幅值无规律可循,导致在引入过程中可能会出现由白噪声信号幅值过低而导致的无法有效抑制模态混合的情况,以及幅值过高致使白噪声残余量过大、高频成分被污染等问题。为此,郑旭等对 EEMD 方法进行了改进,提出了改进的集总平均经验模态分解方法(Modified Ensemble Empirical Mode Decomposition, MEEMD)<sup>[13]</sup>。MEEMD 在加入白噪声时,将分别加入两组正负对称的噪声,然后各自进行 EEMD 分解,这样就可以有效避免由于加入的白噪声幅值过大而导致白噪声残余量过大的问题,同时在分解过程中对分解得到的 IMF 进行二次分解,进而确保分解得到的 IMF 不存在模态分裂问题,而分解得到的 IMF 均为平稳信号。这就解决了非平稳时间序列模式难以识别的根本问题,为设计预测模型建立基础,也为中长期降水量预测提供了新的思路。

根据降水量表现出的复杂特性,结合已有的研究成果,本文提出了中长期降水量预测的混合模型。混合模型可以对多种方法进行综合运用,使各方法发挥各自的优势。在模型中,首先发挥 MEEMD 的优势,将非线性、非平稳的降水量时间序列分解成一组较为平稳的、模态单一的时间序列,这样做可以降低预测方法的预测难度,提高预测精度。接着,根据各序列的特点选择适宜的单项预测方法进行建模。本文主要选择两种预测方法:1)最小二乘法(Least Squares Method, LSM),通过该方法构建回归预测方程对未来值进行预测。相较于 BPNN、SVM 以及极限学习机等机器学习算法,LSM 具有原理简单、方法成熟、计算速度快的优势,但对一些结构复杂的曲线,预测精度不高。2)核极限学习机(Kernel Extreme Learning Machine, KELM),当分解项使用 LSM 时的预测误差过大时,可选择 KELM 作为预测方法进行预测。由于 KELM 的学习性能和泛化能力受到相关参数的影响,为了提高预测方法的性能,有必要对参数进行优化。为此,本文提出

使用改进的果蝇优化算法(Modified Fruit Fly Optimization Algorithm, MFFOA)对 KELM 中的核参数进行寻优,提高预测方法的预测精度,进而提高混合模型整体的预测精度。为了验证混合模型的有效性,本文选用了分布在珠三角地区、粤东地区、粤西地区以及粤北地区的广东省的 7 个观测站的历史降水资料,并对这些地区的年度降水量进行了预测,取得了良好的预测效果,为广东省加强水资源的开发与利用提供了有效的决策支持。

## 2 模型设计

### 2.1 改进的集总平均经验模态分解方法

降水量受大气环流、水文气象要素、自然地理等诸多物理因子的综合作用,是一种弱相关、高复杂度的非线性动力系统,其年际变化并不是以一种固定的周期在运动,而是包含着各种时间尺度的变化和局部波动,这种特性导致中长期降水量的预测难度较大、精度较低。改进的集总平均经验模态分解方法作为一种信号分解方法,适用于分析非线性、非平稳信号,被应用于各类信号分析中<sup>[13-14]</sup>。本文尝试使用 MEEMD 获取降水量序列中的伪周期分量,进而进行建模预测。MEEMD 信号分解的具体步骤参见文献<sup>[13]</sup>。

### 2.2 预测方法

当降水量序列被 MEEMD 分解为  $n$  个不同尺度的 IMF 分量及余项后,就可以为各分解项选择预测方法进行预测。由于各分解项的特性不同,预测方法的适应性和复杂度也有差异,因此各分解项要选择适合其特性的预测方法,这样既可以提高预测模型的运行效率又可以提高预测精度。通过观察分解项时序图可以发现,部分分解项带有明显的趋势特性,是一些有规律可循的曲线。基于此,本文提出在各分解项预测时优先使用 LSM<sup>[15]</sup>。LSM 可以通过现有的数据进行函数拟合,从而使拟合值与真值的误差的平方和最小,同时根据拟合曲线可以简便地求得一些未知的数据。相较于 BPNN、SVM 以及极限学习机等机器学习算法,LSM 具有原理简单、计算速度快的优势。其原理多项式如下:

$$Q(x) = c_0 + c_1x + c_2x^2 + \dots + c_{n-1}x^{n-1} \quad (1)$$

本文使用  $Q(x)$  拟合年度降水量时间序列经 MEEMD 分解得到的  $F(x)$ ,这种拟合的关键是确定  $Q(x)$  中的系数向量  $C = (c_0, c_1, c_2, \dots, c_{n-1})$ ,使得目标函数  $\Delta(c_0, c_1, c_2, \dots, c_{n-1}) = \sum_{i=1}^N (Q(x_i) - F(x_i))^2$  达到最小,  $F(x_i)$  是分解得到的第  $i$  个序列值,  $Q(x_i)$  是多项式  $x=i$  时的值。

在使用 LSM 对分解项进行拟合时,若拟合误差较大,则说明当前分解项不适合使用 LSM 进行预测,这时可以选择核极限学习机作为预测方法进行预测。

极限学习机(Extreme Learning Machine, ELM)<sup>[16]</sup>是一种单隐层前馈神经网络(Single-hidden Layer Feedforward Neural Networks, SLFNs)快速学习方法。其特点是只需指定隐层节点数即可随机生成 SLFNs 隐层的节点参数,并依据正则化最小二乘算法得到网络的输出权值,因此 ELM 能以极快的学习速度获得较好的网络泛化性能。但由于 ELM 在学习过程中其初始权值是随机设定的,没有训练优化

过程,因此会造成算法的性能不稳定。为此,Huang等<sup>[17]</sup>在ELM的基础上提出了基于核函数的极限学习机(Kernel Extreme Learning Machine, KELM),将SVM中核函数的思想运用到ELM中,加快了算法的学习速度,同时提升了算法的精度及泛化能力。

### 2.3 改进的果蝇优化算法

已有研究表明,KELM对参数有一定的敏感性,为了提高KELM方法的预测性能,有必要对其参数进行优化。果蝇优化算法(Fruit Fly Optimization Algorithm, FFOA)<sup>[18]</sup>是由潘文超博士于2011年依据果蝇觅食行为推演出的一种全新的寻求全局优化的新方法。相较于粒子群、鱼群等群体智能优化算法,FFOA具有参数设置少(只有种群大小和最大迭代次数两个参数)、运算速度快且易于程序实现等优点。FFOA的流程参见文献[18]。

根据FFOA学习机制可知,果蝇在寻优过程中只向当前最优的果蝇个体学习,这种学习方式在迭代后期会导致果蝇个体之间具有强烈的趋同性,造成FFOA算法易陷入局部极值,从而导致FFOA的早熟收敛问题,制约了FFOA算法的广泛应用。为了弥补该缺陷,本文对FFOA进行了如下调整:

1)调整搜索机制。将全局最优值搜索过程分解为 $j$ 次独立的局部最优值搜索过程,当搜索次数达到预设值 $N$ 后,退出迭代循环并在 $j$ 次局部最优值中选出全局最优值。通过使用这种搜索机制弥补了经典FFOA算法在陷入局部最优区域后无法跳出的缺陷。为了配合该寻优机制的实施,本文在算法中做了如下改动:

①建立历史局部最优列表 $LocalBestList$ ,该列表用于记录在搜索过程中搜索到的局部最优值及其位置。该列表在搜索迭代完成之后进行一次排序,最终找到全局最优值。

②设定局部最优搜索次数极限值 $M$ 。当果蝇群体在搜索到某最优值 $bestSmell$ 后,又连续搜索 $M$ 次,但仍未找到比 $bestSmell$ 更优的值,则将搜索到的局部最优值 $bestSmell$ 及其位置记录到历史局部最优值列表 $LocalBestList$ ;然后初始化果蝇群的位置,重新计算搜索半径,在全局空间开展搜索工作,这样可以确保果蝇群不会在一个局部区域浪费过多的时间与精力。

2)动态调节搜索半径。在经典FFOA算法中果蝇的搜索半径通常是一个固定值,这种固定值半径的设计机制使得果蝇群在搜索时无法根据需要调整搜索范围。例如,在迭代初期希望在更广阔的区域搜索,此时则需要更大的搜索半径,而在迭代寻优后期,则希望在某个局部区域内展开精细搜索,这时则需要更小的搜索半径。在具体的搜索半径的设计中,本文对每一轮的适应度值的下降梯度与搜索半径之间建立联系,即根据梯度下降幅度来确定搜索半径值,当梯度下降幅度较大时,适当缩小搜索半径,避免由于搜索半径过大、果蝇群分布较为分散而错过最优值;当梯度下降较小时,可适当加大搜索半径。具体的搜索半径的调节方法如下:

①当局部寻优次数 $1 \leq i \leq 3$ 时,有:

$$r(i) = j \times R(j), j \in \{1, 2, \dots, +\infty\} \quad (2)$$

$$R(j) = \begin{cases} R_{init}, & j=1 \\ \|PosLocal(j-1)PosOriginal\|, & j>1 \end{cases} \quad (3)$$

其中, $j$ 表示在总迭代次数内正在进行第几次局部最优值的寻找, $R(j)$ 表示果蝇群的局部寻优初始半径,“ $\times$ ”表示乘积运算, $R_{init}$ 为一常量。 $PosLocal(j-1)$ 表示第 $j-1$ 次迭代得到的局部极值所在位置的坐标, $PosOriginal$ 为果蝇群的初始位置坐标,为了与求绝对值区分,本文使用“ $\| \cdot \|$ ”表示计算两点之间的距离,通过式(3)可以计算出果蝇已经搜索的大致范围,通过式(2)可以使果蝇群在新的局部极值搜索迭代中扩大搜索范围,从而增加搜索到全局最优的可能性。

②当局部寻优迭代次数 $i \geq 4$ 时,根据适应度梯度下降的变化情况,果蝇迭代半径的计算公式如式(4)和式(5)所示。式中 $f(i)$ 表示第 $i$ 次迭代的种群的最优适应度值, $BestPos(i)$ 表示该最优适应度值所处位置的坐标。

当 $f(i-1) - f(i-2) \neq 0$ 并且 $f(i-2) - f(i-3) \neq 0$ 时,有:

$$r(i) = r(i-1) * \frac{\frac{|f(i-2) - f(i-3)|}{\|BestPos(i-2)BestPos(i-3)\|}}{|f(i-1) - f(i-2)|} \quad (4)$$

其中, $i \in \{4, 5, \dots, +\infty\}$ 。

当 $f(i-1) - f(i-2) = 0$ 或者 $f(i-2) - f(i-3) = 0$ 时,有:

$$r(i) = r(i-1), i \in \{4, 5, \dots, +\infty\} \quad (5)$$

改进后的果蝇优化算法的流程如图1所示。

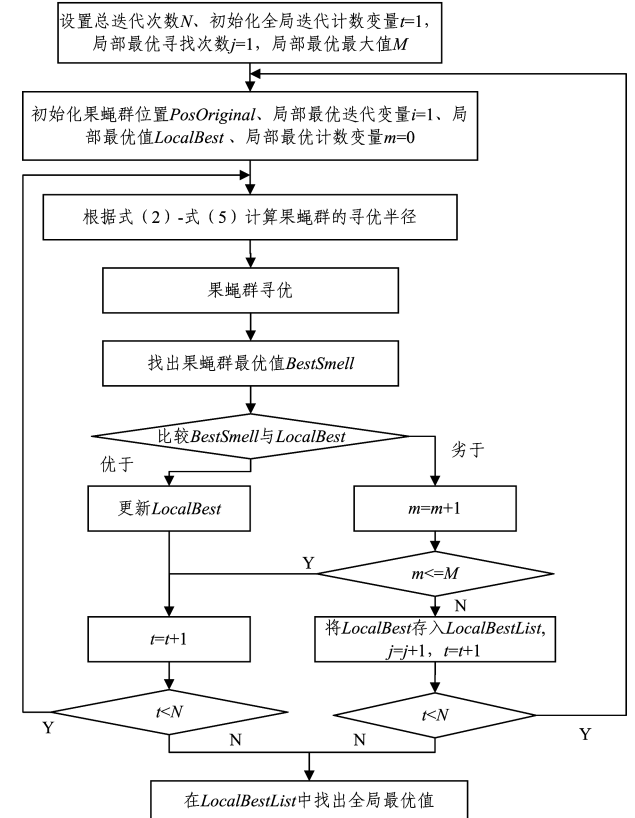


图1 改进后的果蝇优化算法的流程

Fig.1 Flowchart of improved fruit fly optimization algorithm

### 2.4 混合预测模型

鉴于降水量预测的复杂性,本文将前文介绍的几种方法进行混合,提出了降水量混合预测模型,模型的结构如图2所示。

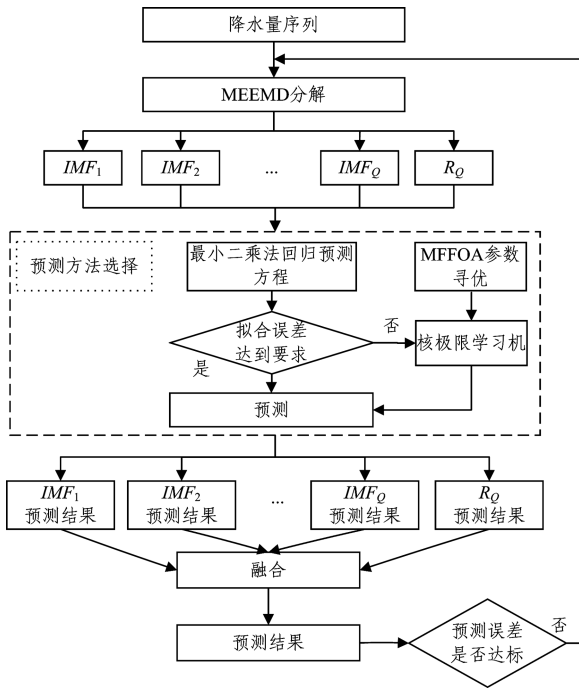


图2 预测模型的结构

Fig.2 Structure of prediction model

建模的具体步骤如下:

1)数据预处理。对收集到的降水量年度时间序列数据进行修复预处理。

2)原始降水量时间序列多尺度分解。采用 MEEMD 对已修复处理的降水量季度值时间序列进行多尺度分解,得到  $Q$  个不同波动频率的本征模态分量( $IMF_1 - IMF_Q$ )和一个残余项  $R$ 。

3)LSM 预测。使用 LSM 对各个分解项数据进行拟合,若拟合误差小于指定值,则确认 LSM 可作为该分解项的预测方法进行预测。若训练数据拟合误差大于指定值,则说明 LSM 不适合当前分解项,需改用 KELM 方法进行预测。

4)KELM 训练数据整理。分解项在采用 KELM 方法进行预测时,首先需要确定 KELM 的输入参数。这里采用相空间重构的方式来确定 KELM 的输入参数,分别使用互信息法和 Cao 方法来确定分解项的延迟时间  $\tau_q$  和嵌入维数  $m_q$ ,并对分解项  $D_q = \{X_i(j), j=1, 2, \dots, N\}$  进行相空间重构,得到相空间域中的数据集  $D'_q = \{X_i(t), Y_i(t), t=1, 2, \dots, M\}$ ,接着对相空间数据集  $D'_q$  进行归一化处理,得到归一化后的数据集  $D''_q = \{X'_i(t), Y'_i(t), t=1, 2, \dots, M\}$ 。

5)KELM 参数的优化。根据  $D''_q$  建立单隐层神经网络,输入数据为  $X'_i(t)$ ,输出数据为  $Y'_i(t)$ ,使用 KELM 方法建立基础预测模型。由于 KELM 的学习性能和泛化能力受到参数的影响,因此本文使用 MFFOA 对 KELM 中的参数进行寻优,寻找最适合当前数据的参数组合。

6)KELM 预测。根据参数优化结果建立基于 KELM 的预测模型,依据归一化之后的相空间数据进行预测,并对预测值进行反归一化。

7)各项预测结果的融合。记第  $q$  个分项在  $t$  时刻的实

测值为  $y_{qt}$  ( $q=1, 2, \dots, Q+1$ ), 预测值为  $\bar{y}_{qt}$  ( $q=1, 2, \dots, Q+1$ ), 则由  $Q+1$  项预测模型构建的混合预测模型在  $t$  时刻的预测值为:

$$\bar{y}(t) = \sum_{q=1}^{Q+1} \bar{y}_{qt} \quad (6)$$

其中,  $Q+1$  表示原始降水量时间序列多尺度分解之后得到的  $Q$  个 IMF 分解项和 1 个残余项  $R$ 。

8)预测结果的评价与反馈。选定预测结果评价指标,并设定相应的指标阈值,当发现预测精度超出设定的阈值时,更新训练集至最新数据,并跳转至步骤 2)重新进行 MEEMD 分解与预测模型的训练。

### 3 模型验证

#### 3.1 研究区概况

广东省地处中国大陆最南部,东邻福建,北接江西、湖南,西连广西,南临南海,珠江口东西两侧分别与香港特别行政区、澳门特别行政区接壤,西南部雷州半岛隔琼州海峡与海南省相望。全境位于北纬  $20^{\circ}13' \sim 25^{\circ}31'$  和东经  $109^{\circ}39' \sim 117^{\circ}19'$  之间,总面积为 17.98 万平方千米。广东省地貌类型复杂多样,有山地、丘陵、台地和平原,其面积分别占全省土地总面积的 33.7%, 24.9%, 14.2% 和 21.7%。地势总体北高南低,北部多为山地和高丘陵,南部则为平原和台地。广东省属于东亚季风区,从北向南分别为中亚热带、南亚热带和热带气候,是全国光、热和水资源较丰富的地区,且雨热同季,降水主要集中在 4—9 月。广东省年平均降水量为 1789.3 毫米,最少年份为 1314.1 毫米,最多年份达 2254.1 毫米。年降水量分布不均,呈多中心分布。

#### 3.2 数据概况

本次研究是在广东省境内 125 个降水量观测站中选择具有代表性的 7 个降水量观测站作为研究对象,这 7 个降水量观测站分别分布在广东省的 4 个区域中,其中粤东地区有汕头,粤北地区有梅州、韶关和河源,珠江三角洲地区有广州,粤西地区有阳江和湛江。研究数据是各降水量观测站从 1951 年—2015 年的年降水数据,序列长度为 65 列,该数据来自于国家气候中心,数据准确且可信。

#### 3.3 预测结果的评价

为了衡量本文模型的稳定性和适应性,选择一些具有代表性的指标对提出的模型进行评价,具体指标有平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE)、均方根误差 (Root Mean Square Error, RMSE) 以及可决系数  $R^2$  (Coefficient of Determination)。与平均百分比误差相比,平均绝对百分比误差能够避免正负相抵的情况,可以更好地反映预测值误差的实际情况。均方根误差对异常大或异常小的误差反应较为灵敏,可以较好地反映模型的预测精度。拟合优度  $R^2$  能够表达预测模型整体的拟合度,  $R^2$  越接近 1, 表明预测值对观测值的拟合程度越好,模型的预测性能也就越好。假设  $T_i$  为实际观察值,  $P_i$  为预测值,各指标的定义如下:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{T_i - P_i}{T_i} \right| \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (T_i - P_i)^2}{N}} \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (T_i - P_i)^2}{\sum_{i=1}^N (T_i - \bar{T})^2} \quad (9)$$

3.4 预测模型的验证过程

3.4.1 降水量时间序列的 MEEMD 分解

广东省各地年度降水量序列具有明显的非线性和非平稳性,这些特性将增加预测的难度。为了更好地对降水量时间序列进行局部分析以获得更高的预测精度,本文利用 MEEMD 对广东省 7 个地区的样本数据进行分解。在分解时,设置白噪音强度为 0.2,集总平均次数为 100。各地区的分解信息如表 1 所列,由表中数据可知,广州、阳江的分解项最多为 16 项;河源、汕头、韶关、梅县分解得到的分解项数次之,均为 15 项;湛江的分解项数最少,为 14 项。

表 1 各地 MEEMD 分解的项数

Table 1 MEEMD decomposition numbers

监测点	广州	河源	汕头	韶关	梅县	阳江	湛江
分解结果	16	15	15	15	15	16	14

为了便于观察各分解项的特性,这里以广州为例给出其年度降水量 MEEMD 分解的结果。由图 3 可知,广州降水量时序数据呈现出明显的多尺度特征,同时这些分解项可以反映出不同影响因素在不同尺度上对降水量的影响。例如,  $IMF_2$  在 1950—1990 年期间,存在着准 3~5 年的波动周期,这与拉尼娜现象的准周期变化较为一致,说明该地区的降水量受到拉尼娜现象的影响;除此之外,  $IMF_6$  也存在类似的准 3~7 年的波动周期,而该波动周期与厄尔尼诺现象的准周期变化较为符合,表明了厄尔尼诺现象与该地区的降雨量变化存在着较为密切的关系,说明该区域降雨量的短期变化受到厄尔尼诺现象的影响。由于当前人类对气候系统的认识还不够深入,因此对于各  $IMF$  分解项的周期性影响因素还未完全探明,而这需要通过未来不断深入的研究来进一步揭示。

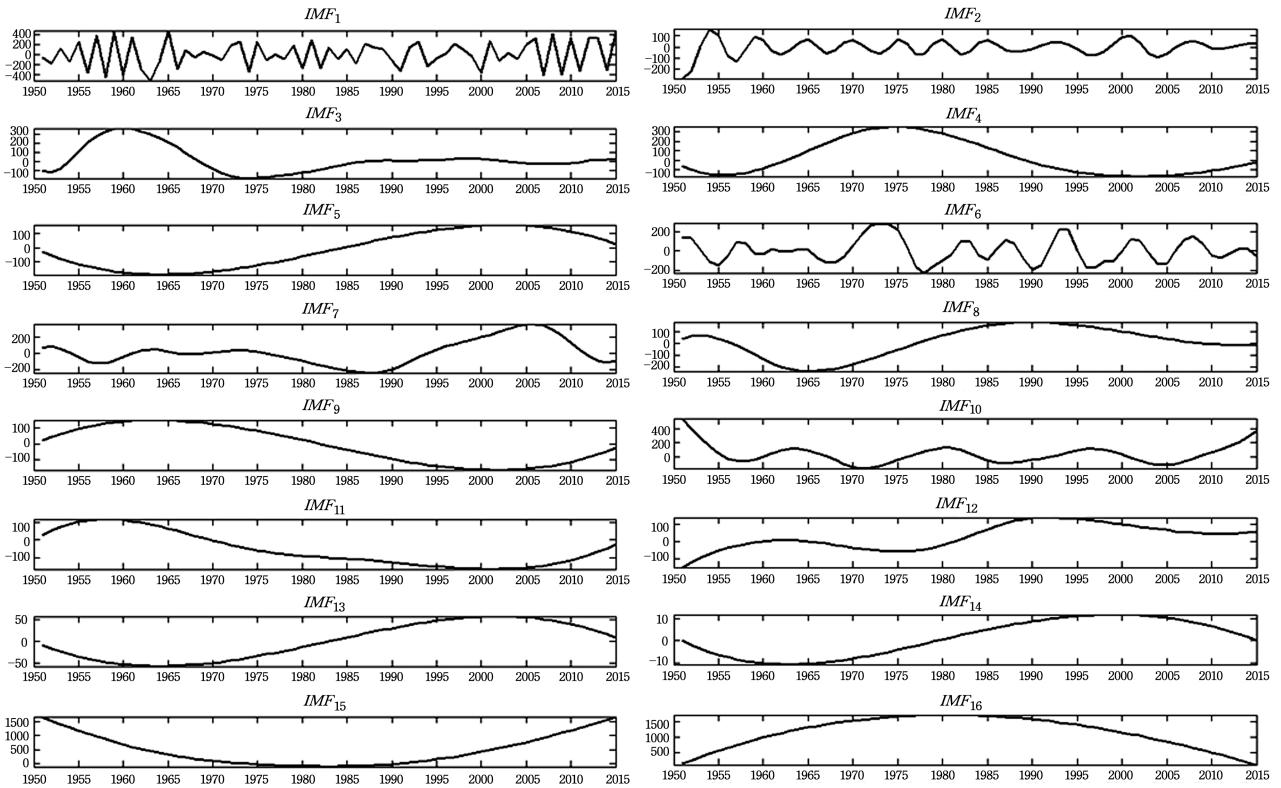


图 3 原始降水量 MEEMD 分解的结果(广州)

Fig. 3 Decomposition results of original precipitation MEEMD(Guangzhou)

3.4.2 LSM 预测

通过对图 3 中各分解项的观察可以发现,部分分解项随时间变化缓慢,具有一定的趋势特性,可以尝试使用 LSM 建立预测多项式进行预测。因此,本文首先使用 LSM 对各分解项进行拟合,并建立回归预测方程。

对各地降雨量的分解项进行 LSM 拟合,并依据均方误差指标进行筛选,然后选择出部分可以使用回归预测方程进行预测的分解项。具体筛选结果如表 2 所列。以广州为例,可以建立回归预测方程的分解项有第 5,9,13 和 16 项,而其余

分解项由于拟合误差较大,无法达到预定标准,将不能采用 LSM 建立回归预测方程进行预测。

表 2 LSM 预测结果

Table 2 LSM prediction results

监测点	广州	河源	汕头	韶关	梅县	阳江	湛江
分解项	5,9,	8,11,	9,12,	6,12,	5,11,	5,9,12,	8,12,
编号	13,16	13,15	15	14	15	14,15	13

3.4.3 KELM 预测

对于未被选择使用 LSM 预测的分解项,将采用 KELM

方法对其进行预测。由于 KELM 是基于单隐层神经网络的机器学习算法,因此在学习之前首先要确定神经网络输入层的各项指标。本文借鉴了混沌时间序列预测中常采用的相空

间重构方法来确定神经网络的输入指标。具体地,在相空间重构中,本文采用互信息法确定延迟时间  $\tau_q$ ,采用 Cao 方法并结合试凑法确定嵌入维数  $m$ ,计算结果如表 3 所列。

表 3 相空间重构信息

Table 3 Phase space reconstruction information

分解项 编号	广州		河源		汕头		韶关		梅县		阳江		湛江	
	延迟 时间	嵌入 维数	延迟 时间	嵌入 维数	延迟 时间	嵌入 维数	延迟 时间	嵌入 维数	延迟 时间	嵌入 维数	延迟 时间	嵌入 维数	延迟 时间	嵌入 维数
1	3	8	2	7	2	12	2	11	2	18	2	18	2	12
2	2	12	2	11	2	13	2	13	2	18	2	11	2	10
3	4	9	2	8	4	11	2	10	5	12	3	10	5	10
4	5	8	8	6	8	6	4	7	6	7	4	8	8	7
5	LSM		3	9	2	15	5	7	LSM		LSM		2	11
6	2	12	2	8	2	11	LSM		2	11	3	11	4	6
7	2	14	8	7	3	6	2	12	4	11	3	12	4	11
8	4	7	LSM		8	6	5	6	3	20	3	11	LSM	
9	LSM		3	8	LSM		3	9	3	7	LSM		3	19
10	3	10	6	8	4	12	3	11	7	7	3	10	6	9
11	2	14	LSM		2	11	2	6	LSM		4	11	6	8
12	2	14	4	6	LSM		LSM		6	7	LSM		LSM	
13	LSM		LSM		4	12	3	10	3	8	5	10	LSM	
14	2	14	2	20	3	14	LSM		3	16	LSM		2	17
15	6	8	LSM		LSM		4	11	LSM		LSM		N/A	
16	LSM		N/A		N/A		N/A		N/A		4	13	N/A	

注:LSM 表示该编号分解项使用的是最小二乘法进行预测,不需要计算延迟时间和嵌入维数。N/A 表示该地区降水量在分解时未生成当前编号分解项,因此也无需计算延迟时间和嵌入维数

依据表 3 中的计算结果,分别对各分解项  $D_q = \{X_i(j), j=1,2,\dots,N\}$  进行相空间重构,得到相空间域中的数据集  $D_q' = \{X_i(t), Y_i(t), t=1,2,\dots,M\}$ 。接着对相空间数据集  $D_q'$  进行归一化处理,从而得到归一化后的数据集  $D_q'' = \{X_i'(t), Y_i'(t), t=1,2,\dots,M\}$ 。依据  $D_q''$  建立单隐层神经网络,输入数据为  $X_i'(t)$ ,输出数据为  $Y_i'(t)$ ,使用 KELM 方法建立基础预测模型。由于正则化系数与核参数对 KELM 的学习性能和泛化能力有一定的影响,因此本文使用改进的 FFOA 对 KELM 中的参数进行寻优,寻找最适合当前数据的参数组合,进而提高预测精度,具体寻优步骤如图 4 所示。

3.4.4 降水量预测结果计算

各分解项通过 LSM 或 KELM 预测可以得到各自的预测值,将这些预测值经累加即可得到降水量预测值。这里依然以广州为例,运算结果如表 4 所列。表中第 1 列为分解项的具体编号,第 2 列为该分解项采用的预测方法,第 3—7 列即各分解项在 2011—2015 年的预测值,表中的最后一行为当前年度各分解项预测值的累加和,即为本模型最终的预测结果。

表 4 分解项预测结果(广州)

Table 4 Forecasting results of each decomposition item (Guangzhou)

分解项 标号	预测 方法	年度				
		2011	2012	2013	2014	2015
1	KELM	-322.09	331.31	331.99	-306.35	484.23
2	KELM	-18.53	-5.95	13.95	29.60	28.19
3	KELM	-10.04	1.23	9.84	13.55	10.96
4	KELM	-95.49	-79.16	-62.33	-42.34	-23.88
5	LSM	103.59	87.41	69.04	48.40	25.41
6	KELM	-76.50	-23.53	21.96	19.61	-54.80
7	KELM	40.56	-35.29	-89.65	-108.88	-84.37
8	KELM	-7.91	-10.64	-11.57	-10.67	-7.98
9	LSM	-101.85	-85.67	-67.46	-47.16	-24.70
10	KELM	97.12	146.48	204.23	280.13	363.83
11	KELM	-99.62	-83.94	-66.81	-46.74	-25.62
12	KELM	42.91	44.07	46.65	50.59	55.84
13	LSM	35.78	30.13	23.75	16.62	8.71
14	KELM	5.44	4.26	2.95	1.51	-0.04
15	KELM	1256.42	1355.60	1441.61	1556.47	1630.69
16	LSM	413.30	334.16	253.22	170.52	86.09
预测结果		1263.09	2010.46	2121.37	1624.85	2472.58

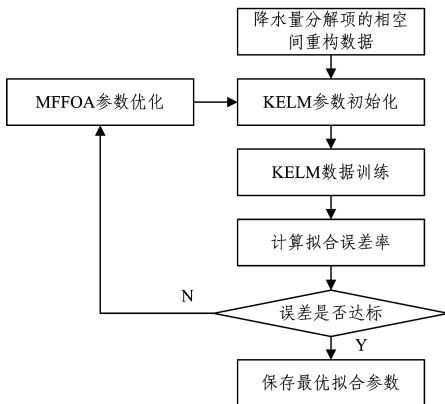


图 4 MFFOA 参数寻优

Fig. 4 Parameter optimization of MFFOA

依据果蝇优化算法的寻优结果建立 KELM 预测模型进行预测,由于预测数据是经过归一化的数据,因此需要对预测结果进行反归一化处理来得到真正的预测值。

3.4.5 预测结果的评价与分析

经过前文介绍的各步骤之后即可得到广东省各地区

2011—2015 年的年度降水量预测值,其实测值与预测值的对比结果如图 5 所示。由图 5 可知,各地降水量的预测值和实际值的拟合情况良好。为了进一步说明模型的预测结果,对各地各年度预测结果的误差百分比进行了统计,统计结果如表 5 所列。通过对比表 5 中的数据可以发现,本文模型最大绝对百分比误差出现在广州地区 2015 年降水量的预测中,误差值为 2.27%,究其原因发现,2015 年广州的降水量创历史新高,最小绝对百分比误差出现在韶关地区 2014 年的降水量预测中,误差为 0.08%。预测模型对各地的预测表现也存在一定差异,其中汕头的平均绝对百分比误差 MAPE 最大,误差值为 1.39%;韶关的平均绝对百分比误差 MAPE 最小,误差值为 0.38%。总体来看,本文预测模型预报合格率为 100%(以绝对百分比误差小于 20%为合格标准<sup>[19]</sup>)。由此可知,本文提出的混合模型是一种合适宜预测年度降水量的新模型。

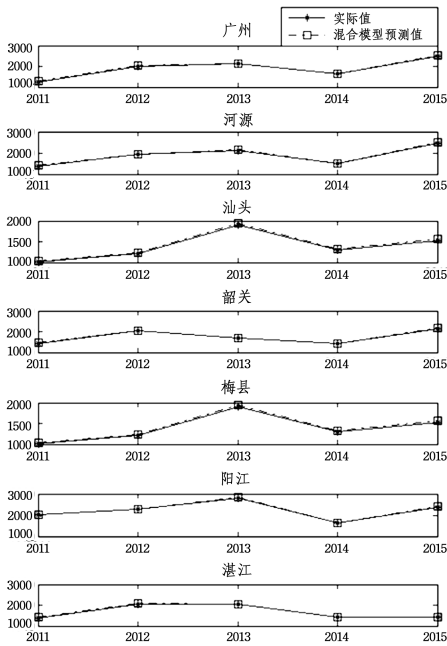


图 5 降水量预测值与实测值的比较

Fig. 5 Comparison between forecasting values and actual values of rainfall

表 5 预测误差统计

Table 5 Forecasting error statistics

	广州	河源	汕头	韶关	梅县	阳江	湛江
2011	1.48	1.37	1.15	0.35	0.72	0.49	0.53
2012	0.86	0.83	1.15	0.29	0.46	0.71	0.51
2013	1.51	1.28	1.39	0.52	0.91	0.91	0.67
2014	0.56	0.45	1.02	<b>0.08</b>	0.34	0.29	0.36
2015	<b>2.27</b>	1.49	2.23	0.68	1.51	1.71	1.02
MAPE	1.34	1.09	1.39	0.38	0.79	0.82	0.62

(单位:%)

为了进一步说明本文模型的有效性,将本文模型与自回归滑动平均模型(Auto-Regressive and Moving Average Model, ARMA)以及 KELM 预测模型的预测结果进行进一步对比,各模型的预测结果如图 6 所示。由图 6 可以直观地看出,

本文模型的预测结果更加接近真实值,也就是说混合模型具有更好的预测精度。为更加全面地评价模型性能,表 6 给出了模型性能的统计值 MAPE、RMSE、 $R^2$  以及预报合格率。从表 6 可以看出,混合模型在这 4 项性能评价指标上明显优于其他两种模型。例如,河源站点混合模型的 MAPE, RMSE 由 ARMA 模型的 18.03%, 353.26 以及 KELM 模型的 9.26%, 193.22 降低到 1.09%, 23.73; 并且  $R^2$  和预报成功率分别由 ARMA 模型的 0.255, 60.00% 以及 KELM 模型的 0.777, 80.00% 显著提高到 0.997, 100.00%。由此可以看出,混合预测模型具有更加准确的时序数据预测能力。纵观表 6 中 7 个观测站点(地区)的性能评价指标数据,本文模型均能降低 MAPE 和 RMSE, 提高  $R^2$  和预报合格率值, 这表明本文模型具有较强的泛化能力, 能够适应各地降水量的模式变化, 保持较高的预测精度。总之, 相较于 ARMA 和 KELM, 本文模型能够准确地预测各地年度降水量的变化趋势, 可为地区开展水资源的优化管理提供有效的指导。

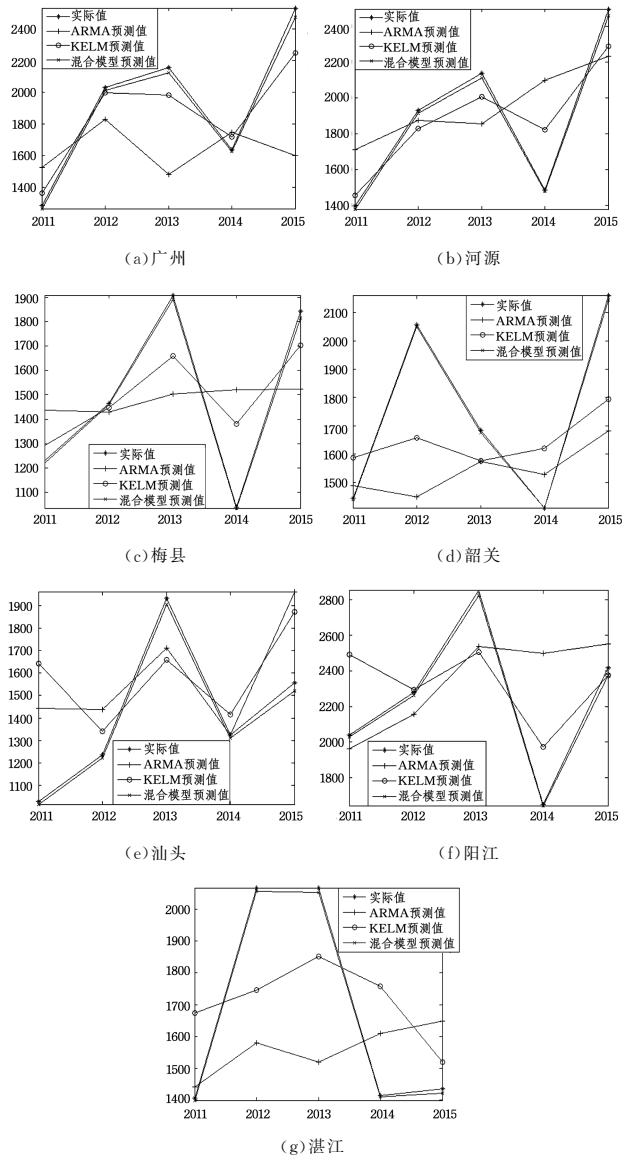


图 6 多模型对年度降水量的预测结果时序图

Fig. 6 Time-series plots for predicting annual precipitation by multiple models

表6 模型预测性能比较

Table 6 Comparison of predictive performance of models

监测点	模型	MAPE/%	RMSE	R <sup>2</sup>	预报合格率/%
广州	ARMA	20.80	536.47	-0.552	60.00
	KELM	6.40	157.10	0.867	100.00
	混合模型	<b>1.34</b>	<b>31.97</b>	<b>0.995</b>	<b>100.00</b>
河源	ARMA	18.03	353.26	0.255	60.00
	KELM	9.26	193.22	0.777	80.00
	混合模型	<b>1.09</b>	<b>23.73</b>	<b>0.997</b>	<b>100.00</b>
汕头	ARMA	18.95	292.88	0.109	60.00
	KELM	22.07	339.75	-0.198	60.00
	混合模型	<b>1.39</b>	<b>22.14</b>	<b>0.995</b>	<b>100.00</b>
韶关	ARMA	13.93	353.50	-0.315	60.00
	KELM	13.54	272.08	0.221	100.00
	混合模型	<b>0.38</b>	<b>8.41</b>	<b>0.999</b>	<b>100.00</b>
梅县	ARMA	20.81	329.44	0.048	60.00
	KELM	11.96	201.52	0.644	80.00
	混合模型	<b>0.79</b>	<b>15.61</b>	<b>0.998</b>	<b>100.00</b>
阳江	ARMA	15.47	415.41	-0.084	80.00
	KELM	11.28	293.87	0.457	80.00
	混合模型	<b>0.82</b>	<b>23.47</b>	<b>0.997</b>	<b>100.00</b>
湛江	ARMA	16.20	352.18	-0.233	60.00
	KELM	14.99	262.44	0.315	80.00
	混合模型	<b>0.62</b>	<b>10.96</b>	<b>0.999</b>	<b>100.00</b>

**结束语** 降水量时间序列具有复杂性、层次性的特征,通过使用 MEEMD 方法可以将看似无规律的降水量时序数据分解为多个分解项,能够有效减少各分解项之间的干扰,降低预测的复杂程度。在分析各分解项序列时发现,部分分解项背后的成因已经有所探明,如拉尼娜现象和厄尔尼诺现象,但仍有部分分解项的成因依然无法得到有效解释,而这需要通过后续的深入研究来进一步揭示。

通过 MEEMD 方法分解得到的分解项的特性各有不同,而预测方法的适应性也各不相同,因此需要为每个分解项选用适合的预测方法,这样才能够提高整体的预测精度和预测效率。

本文模型是将年度降水量时序数据通过 MEEMD 分解后,再对每个分解项进行预测,与传统的 ARMA 以及 KELM 模型相比,本文模型增加了一定的计算代价。然而年度降水量数据样本较少,并且 MEEMD 分解计算简单,因此总体上本文模型增加的运行时间有限,不会对本文模型的实践应用产生影响。

从实验结果可以看出,本文模型对于年度降水量预测的精度较高,尤其是在年度降水量变化幅度较大时,本文模型依然能够表现出较高的准确性,因此可以认为年度降水量数据经 MEEMD 分解后可以较为准确地识别出降水量数据中的层次特征,有助于研究者选择更适合的预测方法来提高整体的预测精度。在与 ARMA 以及 KELM 预测模型比较中,本文模型表现出了良好的预测精度和稳定性,同时本文模型的设计较为简单,具有一定的实用性。本文的研究结果不仅能有效应用于地区年度降水量的时序数据预测,同时也可用于其他时间维度降水量的时序预测研究,可为相关部门开展水资源工程、管理和计划提供理论支持和决策依据。

## 参考文献

[1] ALTUNKAYNAK A, NIGUSSIE T A. Prediction of daily rainfall by a hybrid wavelet-season-neuro technique[J]. Journal of Hydrology, 2015, 529(10): 287-301.

- [2] SURUSSAVADEE C. Evaluation of High-Resolution Tropical Weather Forecasts Using Satellite Passive Millimeter-Wave Observations[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(5): 2780-2787.
- [3] NAGESWARARAO M M, MOHANTY U C, OSURI K K, et al. Prediction of winter precipitation over northwest India using ocean heat fluxes[J]. Climate Dynamics, 2016, 47(7): 2253-2271.
- [4] KISI O, SANIKHANI H. Prediction of long-term monthly precipitation using several soft computing methods without climatic data[J]. International Journal of Climatology, 2016, 35(14): 4139-4150.
- [5] AUERBACH D A, EASTON Z M, WALTER M T, et al. Fuka Evaluating weather observations and the Climate Forecast System Reanalysis as inputs for hydrologic modelling in the tropics[J]. Hydrological Processes, 2016, 30(19): 3466-3477.
- [6] HU F L, WANG L Q, ZUO R T, et al. Extra-seasonal Predicting Tests and Analyses of Several Statistical Forecasting Methods on Precipitation over Nanjing in 1998[J]. Climatic and Environmental Research, 2017, 22(1): 23-34. (in Chinese)  
胡凤良, 王丽琼, 左瑞亭, 等. 几种统计预测方法对 1998 年南京降水的跨季节预测试验及分析[J]. 气候与环境研究, 2017, 22(1): 23-34.
- [7] HUANG W, LI X Z, ZHAO J, et al. A precipitation forecasting method for a river basin based on naive Bayes algorithm[J]. Advances in Science and Technology of Water Resources, 2016, 36(4): 65-69, 79. (in Chinese)  
黄炜, 李雪真, 赵嘉, 等. 基于朴素贝叶斯算法的流域降水预测方法[J]. 水利水电科技进展, 2016, 36(4): 65-69, 79.
- [8] HUANG C Y, HAN Z W, CHANG J X, et al. Prediction Analysis of Rainfall Series Based on EEMD and GRNN Model[J]. Yellow River, 2017, 39(5): 26-28. (in Chinese)  
黄春艳, 韩志伟, 畅建霞, 等. 基于 EEMD 和 GRNN 的降水量序列预测研究[J]. 人民黄河, 2017, 39(5): 26-28.
- [9] BAI Y, CHEN Z Q, XIE J J, et al. Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models[J]. Journal of Hydrology, 2016, 532: 193-206.
- [10] XUE C F, HOU W, ZHAO J H, et al. The application of ensemble empirical mode decomposition method in multiscale analysis of region precipitation and its response to the climate change[J]. Acta Physica Sinica, 2013, 62(10): 504-511. (in Chinese)  
薛春芳, 侯威, 赵俊虎, 等. 集合经验模态分解在区域降水变化多尺度分析及气候变化响应研究中的应用[J]. 物理学报, 2013, 62(10): 504-511.
- [11] GU X, ZHANG Y M, HUANG D W. Temporal and Spatial Characteristics of Soil -Air Temperature Difference (Ts-Ta) in Southeast Guizhou Last 50 Years [J]. Chinese Journal of Agrometeorology, 2012, 33(1): 71-77, 85. (in Chinese)  
顾欣, 张艳梅, 黄大卫. 近 50 年来黔东南地区浅层地气温差的时空特征分析[J]. 中国农业气象, 2012, 33(1): 71-77, 85.
- [12] WU Z, HUANG N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method[J]. Advances in Adaptive Data Analysis, 2009, 1(1): 1-41.

聚类算法对 FAsT-Match 算法进行改进,对 FAsT-Match 算法得到的变换矩阵进行聚类,从而实现多靶位的定位。精准、快速地定位多靶位为激光模拟设计系统的后续设计提供了极大的便利。

1) 本文所选用的参数  $\delta=0.3, \epsilon=0.25$ , 在实际应用中可根据环境如摄像头远近、模板图像大小等来适当调整参数;

2) 当  $\epsilon>0.3$ , 采样点数越多时, 算法的运行时间趋于相同, 但此时的定位精度降低;

3) 目前设计的激光模拟射击系统靶位与摄像头是逐一对应的, 将本文方法应用到系统中可减少模拟射击系统中摄像头的数量, 从而节约摄像头成本;

4) 精准、快速地定位多靶位为激光模拟设计系统的后续设计提供了极大的便利。

## 参 考 文 献

- [1] LIN Y G, HE X H, ZHANG S J. Real-time Multi-targets Tracking Based on Region Template Matching [J]. Computer Simulation, 2011, 28(1): 277-280. (in Chinese)  
林月贡, 何小海, 张生军. 基于区域模板匹配的多目标实时跟踪[J]. 计算机仿真, 2011, 28(1): 277-280.
- [2] XIN N F, QIAN Q C, PENG H D, et al. Design of automatic target-scoring system of shooting game based on computer vision [C]//Proceedings of the IEEE International Conference on Automation and Logistics. Shenyang, China, 2009.
- [3] ALI F, MANSOOR A B. Computer vision based automatic scoring of shooting targets [C]// Multitopic Conference, 2008 (INMIC 2008), IEEE International. IEEE, 2008: 515-519.
- [4] LEE W C, CHEN C H. A fast template matching method for rotation invariance using two-stage process [C] // Proceeding of IEEE Conference on Intelligent Information Hiding and Multimedia Signal Processing. Kyoto, Japan, 2009: 9-12.
- [5] LV J, WANG B, GAO H M, et al. Separate of Multi objects in Image Recognition by Local Features [J]. Acta Photonica Sinica, 2008, 37(8): 1708-1712. (in Chinese)  
吕冀, 汪渤, 高洪民, 等. 图像局部特征识别中的多目标分离[J]. 光子学报, 2008, 37(8): 1708-1712.
- [6] GAO W, ZHU M, HE B G, et al. Overview of target tracking technology [J]. Chinese Optics, 2014, 7(3): 366-375. (in Chinese)  
高文, 朱明, 贺柏根, 等. 目标跟踪技术综述[J]. 中国光学, 2014, 7(3): 366-375.
- [7] KORMAN S, REICHMAN D, TSUR G, et al. FAsT-Match: Fast Affine Template Matching [J]. International Journal of Computer Vision, 2017, 121(1): 111-125.
- [8] LEI L, LI Z Y, SU Y. Multi-target Tracking Based on Fusion of ROI Feature Matching [J]. Journal of Image and Graphics, 2008, 13(3): 580-585. (in Chinese)  
雷琳, 李智勇, 粟毅. 基于 ROI 特征匹配融合的图像多目标跟踪算法[J]. 中国图象图形学报, 2008, 13(3): 580-585.
- [9] ZOU H, HUANG F Z. Infrared Image Segmentation for Electrical Equipment Based on FAsT-Match Algorithm [J]. Infrared Technology, 2016, 38(1): 21-27. (in Chinese)  
邹辉, 黄福珍. 基于 FAsT-Match 算法的电力设备红外图像分割[J]. 红外技术, 2016, 38(1): 21-27.
- [10] 冈萨, 雷斯. 数字图像处理(第二版) [M]. 北京: 电子工业出版社, 2007.
- [11] ZHAO L L, GENG G H, LI K, et al. Images matching algorithm based on SURF and fast approximate nearest neighbor search [J]. Application Research of Computers, 2013, 30(3): 921-923. (in Chinese)  
赵璐璐, 耿国华, 李康, 等. 基于 SURF 和快速近似最近邻搜索的图像匹配算法[J]. 计算机应用研究, 2013, 30(3): 921-923.
- [12] HOU Y M, DI J M. Application of Improved Scale Invariant Feature Transform Accurate Image Matching in Target Positioning of Electric Power Equipment [J]. Proceedings of the CSEE, 2012, 32(10): 134-139. (in Chinese)  
侯一民, 邸建铭. 改进的尺度不变特征转换精确图像匹配在电力设备目标定位中的应用[J]. 中国电机工程学报, 2012, 32(10): 134-139.
- [13] LI G, FENG Y. Moving object detection based on SFT features matching and k-means clustering [J]. Journal of Computer Applications, 2012, 32(10): 2824-2826. (in Chinese)  
李广, 冯燕. 基于 SIFT 特征匹配与 K-均值聚类的运动目标检测[J]. 计算机应用, 2012, 32(10): 2824-2826.
- [14] ZHENG X, HAO Z Y, LU Z G, et al. Separation of piston-slap and combustion shock excitations via MEEMD method [J]. Journal of Vibration and Shock, 2012, 31(6): 109-113. (in Chinese)  
郑旭, 郝志勇, 卢兆刚, 等. 基于 MEEMD 的内燃机机体活塞敲击激励与燃烧爆发激励分离研究[J]. 振动与冲击, 2012, 31(6): 109-113.
- [15] NIU C, LI X H, YI S H, et al. Forecasting Model of Geomagnetic Variation Field Based on Modified Ensemble Empirical Mode Decomposition-Sample Entropy-east Square support Vector Machine [J]. Geomatics and Information Science of Wuhan University, 2014, 39(5): 626-630. (in Chinese)  
牛超, 李夕海, 易世华, 等. 地磁变化场的 MEEMD-样本熵-LSS-VM 预测模型[J]. 武汉大学学报(信息科学版), 2014, 39(5): 626-630.
- [16] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: Theory and applications [J]. Neurocomputing, 2006, 70(1): 489-501.
- [17] HUANG G B, ZHOU H M, DING X J, et al. Extreme Learning Machine for Regression and Multiclass Classification [J]. IEEE Transactions on Systems, Man, and Cybernetics, 2012, 42(2): 513-529.
- [18] PAN W C. A New Fruit Fly Optimization Algorithm: Taking the Financial Distress Model as An Example [J]. Knowledge-Based Systems, 2012, 26(2): 69-74.
- [19] XING Z X, GUO H, FU Q. Analysis of Influencing Factors of Rainfall in Irrigation Area and Combining Rainfall Forecasting [J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(8): 97-103. (in Chinese)  
邢贞相, 郭皓, 付强. 灌区降水量影响因子分析与集合降水预测[J]. 农业机械学报, 2015, 46(8): 97-103.
- [20] thod [J]. Hydrological Processes, 2015, 14(6): 1003-1016.

(上接第 278 页)