

杜良敏,柯宗建,刘长征,等. 2016. 基于聚类分区的中国夏季降水预测模型. 气象, 42(1):89-96.

基于聚类分区的中国夏季降水预测模型^{* 1}

杜良敏¹ 柯宗建² 刘长征² 肖 莺¹ 刘绿柳²

1 武汉区域气候中心, 武汉 430074

2 国家气候中心, 北京 100081

提 要: 文章基于近邻传播客观聚类方法对中国夏季降水进行了气候分区, 以中国不同分区的夏季降水为预测对象, 使用前期的海温和海平面气压场为预测因子, 利用图像标签算法提取高相关封闭区域的预测因子信息, 结合最小二乘回归法建立预测模型。采用 P_s 评分、距平符号一致率和距平相关系数三种评分方法检验了该预测模型, 比较了四种不同的因子配置方案的预测能力。研究表明, 利用冬春季海温的演变特征结合海平面气压的年际变化为预测因子的分区预测模型效果较好, 在 1982—2009 年期间的平均交叉检验平均 P_s 得分为 81.4, 距平符号一致率为 63%, 距平相关系数为 0.35, 2010—2014 期间的独立样本预测检验的平均评分分别为 77.1, 58% 和 0.19, 且逐年回报效果较为稳定, 表明该方法对中国夏季降水有较好的预测效果。研究结果显示, 该预测模型能较好地预测出 2014 年中国夏季降水南多北少的分布特征。

关键词: 近邻传播聚类, 分区, 夏季降水预测, 图形标签算法

中图分类号: P467

文献标志码: A

doi: 10.7519/j.issn.1000-0526.2016.01.011

Summer Precipitation Prediction Models Based on the Clustering Regionalization in China

DU Liangmin¹ KE Zongjian² LIU Changzheng² XIAO Ying¹ LIU Lüliu²

1 Wuhan Regional Climate Centre, Wuhan 430074

2 National Climate Centre, Beijing 100081

Abstract: Based on affinity propagation clustering method, summer precipitation is regionalized over China. Summer precipitation in different regions serves as predicted objects, and preceding sea surface temperature and sea level pressure are selected to be predictors. By methods of image labeling algorithm, consecutive high correlation areas are extracted to determine the predictors. Least squares regression method is used to construct a model to predict summer precipitation in different regions. Different scoring methods including P_s , the anomaly sign sameness rate and anomaly correlation coefficient are used to validate the skills of prediction model in four different factor combination schemes. The results show that the model performs best when sea surface temperature and sea level pressure in winter and spring are together considered as factors. The averaged P_s score of cross validation is 81.4 from 1982 to 2009, with the anomaly sameness rate 63% and anomaly correlation coefficient 0.35. The retrospective forecast verification shows that the averaged scores from 2010 to 2014 are 77.1, 58% and 0.19, respectively. The reforecast skills are relatively stable every year, which means that the method has a good ability to predict summer precipitation in China. Moreover, it succeeds to predict the spatial characteristic of southern flood and northern

* 公益性行业(气象)科研专项(GYHY201306024, GYHY201306033 和 GYHY201406022)、国家重大科学研究计划(2012CB955902)和国家自然科学基金项目(41005051 和 41205039)共同资助

2015 年 1 月 13 日收稿; 2015 年 9 月 9 日收修定稿

第一作者: 杜良敏, 主要从事短期气候预测和信息技术应用研究. Email: duliangmin@outlook.com

通信作者: 柯宗建, 主要从事短期气候预测研究. Email: kezej@cma.gov.cn

drought in China in summer 2014.

Key words: affinity propagation clustering, regionalization, summer precipitation prediction, image labeling algorithm

引 言

中国是个自然灾害频发的国家,气象灾害造成的损失占全部自然灾害损失的 70% 以上。气象灾害具有种类多、发生频率高、影响范围广、持续时间长的特点,且呈经济越发达,损失越大的趋势,因此短期气候预测对防御和减轻自然灾害促进国民经济和社会发展有重要的意义。目前,统计方法在短期气候预测中仍占有非常重要的位置,在充分考虑天气、气候的演变规律和物理机制的基础上,设计统计预测模式或方案,在短期气候预测领域仍发挥着重要的作用。20 世纪 80 年代以来,由于实际工作的需要以及数值计算能力的迅猛发展,基于多个自变量(预测因子)与因变量(预测量)的回归分析、岭回归分析、主成分回归分析、最优子集回归等方法被广泛地应用到气候预测建模中(范丽军等,2007;柯宗建等,2009;杜良敏等,2011)。

中国东西跨度大、南北距离长,具有从西部青藏高原到东部平原的复杂地形和多种下垫面条件,气候呈现出复杂的区域特征。因而,在适当的气候分区基础上,以区域气候特征作为预测对象,有利于获取大尺度影响因子,提高区域气候可预报性(顾伟宗等,2012)。

目前针对中国不同区域夏季降水预测开展了若干方法的应用研究,但是基于客观聚类的分区气候要素特征作为预测对象开展的研究工作仍不多。近几十年来在气候分区方面曾有过多方面研究,在综合气候区划方面,利用最大可能蒸发量(热量)和湿润度指数(降水和蒸发)对中国进行了二级区划,将中国划分为 5 个气候带和 10 余个气候区(张家诚,1991)。利用年积温、候、月平均、极端最低气温以及干燥指数 K 将中国分为 8 个一级区、32 个二级区(气候省)和若干个三级区,同时在区域性的气象要素分区方面也有一些相关的研究工作(马京津等,2012)。在短期气候预测业务中,为了反映各地区的降水情况,根据地理位置和气候特征,国家气候中心按照地理分布等方式,将全国分为 15 个分区(陈兴芳等,2000),每个区域内选取若干个代表站,由各个

代表站的月降水资料计算降水指数作为预测分析的对象。上述分区有从降水(或干湿)的一致性角度考虑的,也有从大气环流的一致性角度考虑的,还有的从农业气候资源的角度来考虑。总的来说,大的综合气候分区只是部分隐含小区域的降水特征。在各类更密集的测站资料条件下,进一步细化降水分区特征尤为必要。

从气候预测的角度而言,基于气候分区的预测对象其年际变化及气候变率相对一致,而目前的气候业务中习惯用行政区划或约定俗成的固定区域代表站来做分析和预测,这忽略了行政区划或固定区域内不同站点气候年际变化可能存在的不同,与气候分区的划分会存在不匹配的情况。因此,基于中国夏季降水观测资料,利用聚类簇内关联度高,聚类簇外关联度低的客观聚类分区方案开展降水气候分区研究,并以夏季气候分区降水为对象发展中国区域夏季降水的预测模型是本文的研究目的。

1 资 料

本文的降水资料采用国家气象信息中心提供的全国 326 站 1981—2014 年的夏季(6—8 月)降水资料,剔除历史资料不全的站点,实际站点 317 站。海温场资料采用 NOAA 最优插值海表面温度第二版本的逐月平均资料,网格距为 $1^\circ \times 1^\circ$ (Reynolds et al, 2002),起始时间为 1980 年 12 月至 2014 年 5 月。海平面气压场采用 NCAR/NCEP 再分析资料,网格距为 $2.5^\circ \times 2.5^\circ$ (Kalnay et al, 1996),起始时间为 1981 年 1 月至 2014 年 5 月。

2 中国夏季降水近邻传播聚类

聚类分析在机器学习的分支非监督学习中应用非常广泛,已经成为大数据应用领域最热门的研究内容之一。有多种方法和手段应用于聚类分析中,在气象学领域主要应用的聚类方法有 REOF 聚类、K-Mean 聚类、模糊聚类、层次聚类、谱聚类、Ward 聚类、聚类统计检验法等(刘杨等, 2012;徐国昌等, 1989;赵汉光等, 1993;江志红等, 1994;李维京等,

1999;丁裕国等,2007;金荣花等,2007;毛伟峰等,2008;赵超等,2008;Chen et al,2009;刘伟东等,2014;韩微等,2015),这些方法广泛地应用于气候分区区划、诊断分析和预报预测工作中。

但是在实际的气象要素的分区分析工作中,一些研究采用了主观分区的方式(郭艳君等,2014;张小曳,2014)。在分析了各种聚类方法的主要特点基础上,本文采用了一种较新的客观聚类方法,近邻传播聚类(Affinity propagation clustering;Frey et al,2007)。该算法根据 N 个数据点之间的相似度进行聚类,这些相似度可以是对称的,即两个数据点互相之间的相似度一样(如欧氏距离、余弦相似度或其他度量方式等);也可以是不对称的,即两个数据点互相之间的相似度不等。这些相似度组成 $N \times N$ 的相似度矩阵 S (其中 N 为有 N 个数据点)。近邻传播算法可以不事先指定聚类数目,也可以指定输出聚类数目,它将所有的数据点都作为潜在的聚类中心,称之为 exemplar。以 S 矩阵的对角线上的数值 $s(k, k)$ 作为 k 点能否成为聚类中心的评判标准,这意味着该值越大,这个点成为聚类中心的可能性也就越大,这个值又称作参考度 p (preference)。聚类的数量受到参考度 p 的影响,如果认为每个数据点都有可能作为聚类中心,那么 p 就应取相同的值。如果取输入的相似度的均值作为 p 的值,得到聚类数量是中等的。如果取最小值,得到类数较少的聚类,可以通过调整参考度 p 来得出相应的类值,目前该聚类方法已经得到了较为广泛的应用(王开军等 2007,肖宇等 2008)。

由于本文中的聚类分区主要用于气候预测,采用客观聚类时既要考虑站点之间的相关性、也要考虑到站点的距平符号一致率,也要剔除掉遥相关的站点,综合考虑以上的三个特点,本文构造了考虑以上三个要素的距离函数。一般用于聚类时两者之间的距离越小越近,越容易聚集,式(1)中的 $R(i, j)$ 代表两个站点之间的相关系数距离,当相关系数为正且显著性水平达到 0.01 时,其相关系数距离为 0,未达到时为 1 减去其相关系数,这样距离越大。式(2)代表距平符号同号率距离,当站点之间的历史降水的距平符号越一致时,其距离越小。式(3)代表地理上的大圆距离,当大圆距离在 300 km 以内时,将此距离设置为 0,以避免干扰相关性高、符号一致率高且相邻的站点,当大圆距离大于等于 300 km 时,该距离就等于大圆距离除以一个系数 $Coef$,使其量

级与式(1)和式(2)中的距离相匹配,该系数在此处取 500,此处大圆距离阈值之所以取 300 km,主要考虑中国中东部相邻省会城市之间的大圆距离在 200~300 km 左右,使用这个阈值容易使空间上相邻的站点聚集在一起,空间距离远的站点尽可能分开。公式(4)代表构造的两个站点的相关距离为式(1)~(3)的三者之和。利用式(4)对全部 317 站的分别建立任意两点之间的距离,构造 317×317 大小的相似度矩阵 S 。

$$R(i, j) = \begin{cases} 0, & pval(i, j) < 0.01 \text{ 和 } PearsonR(i, j) > 0 \\ 1 - PearsonR(i, j), & pval(i, j) \geq 0.01 \end{cases} \quad (1)$$

$$A(i, j) = 1 - Anomaly \text{ Sign Rate}(i, j) \quad (2)$$

$$G(i, j) = \begin{cases} 0, & \text{Great Circle Distance}(i, j) < 300 \text{ km} \\ \text{Great Circle Distance}(i, j) / Coef, & \text{Great Circle Distance}(i, j) \geq 300 \text{ km} \end{cases} \quad (3)$$

$$s(i, j) = R(i, j) + A(i, j) + G(i, j) \quad (4)$$

运用近邻传播算法搜索聚类空间能够输出一系列具有不同聚类数目的聚类结果,对聚类结果进行分析并确定最佳聚类数这是一个难点,在综合考虑前期针对夏季降水分区所做的工作(陈兴芳等,2000;Chen et al,2009),在本文中为分析方便将聚类数定为 15 个区域,采用前述的相似度量距离,采用近邻传播聚类最后得出距离分区的结果(图 1),分析该图可以得出在空间上形成了各个分区内站点相互靠近的结构,避免了跨区聚集为同一分区的现象。

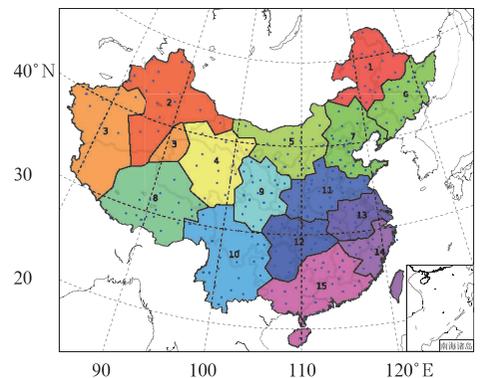


图 1 中国夏季降水近邻传播聚类分区
Fig. 1 Affinity propagation clustering regionalization of summer precipitation in China

3 预测因子的选取

预测因子的选取是建立统计模型非常重要的环节,因子的选择是否合适对预测模型的结果有很大影响。对于不同区域的预测对象来说,其预测因子可能存在差异。本文参考了相关文献(柯宗建等,2009;顾伟宗等,2009;刘绿柳等,2011)在最优子集降尺度方法中的因子提取方式,并加以改进,提出了一种较新的格点场因子提取的方法,该方法不固定提取因子区域范围大小,综合考虑相关的置信区间和范围来选取因子。首先,建立每个聚类分区观测站点均值的历史序列,再计算其与再分析要素场资料之间的相关系数。然后在全球范围内搜索相关系数达到 0.01~0.1 显著性水平的区域,并保证连通区域面积大于指定的经纬度网格点面积,在本文中经纬度网格点取 100。以海温场为例,目前常用海温指数资料有 ENSO 海温指数和印度洋海温指数,如 Nino3、Nino3.4、SIOD 等,大多为 100~900 个经纬度格点。将所有满足以上条件的封闭区域所在格点挑选出来,如果达到上述标准,则只选取信度和面积达到上述条件的前 3 个因子作为预测因子集,如果未达到标准,则不予采用。这样既保证了预测因子与预测量之间的高相关性,又保证了它们之间在一定程度上的物理联系,具体计算步骤如图 2 所示。

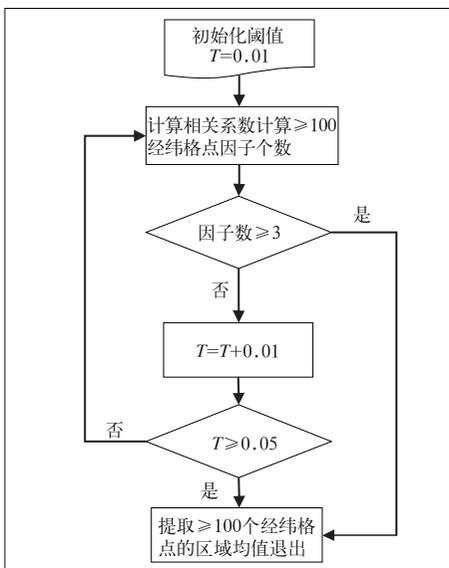


图 2 因子场封闭区域高相关系数的提取方式

Fig. 2 Extraction method for regional factor with high correlation coefficient

4 预测与检验

海洋与大气相互作用是气候研究领域的一个核心内容。大量的事实和理论研究表明,海洋几乎在所有的尺度上的气候变化中起着重要的作用(蔡榕硕等,2012;郭玲等,2012;袁媛等,2012;徐志清和范可,2012)。海洋的年代际和年际变化特征是季节预测的一个重要因素。由于我国夏季降水趋势预测业务会商于每年的 3 月底举行,热带海洋和大气的变化是汛期预测关注的重点之一(陈丽娟等,2013;柯宗建等,2014)。因此,选取前期海温和海平面气压场作为预测因子场,目前在气候预测领域对年际增量作为预报对象有一定的应用(王会军等,2010;2012;肖科丽等,2015),所以本文在因子的选取方式上不仅考虑了年际海温的变化,也考虑了冬春季海温的变化幅度对后期的影响,所以在因子组合上考虑了三种方案,对第三种方案的因子提取方式不变,将预测对象调整为分区内的单站预测对象,形成第四种方案,总共得出四种预测方案,具体的配置方案见表 1。

利用最小二乘回归法,对选取的因子做回归建模。分别采用交叉检验(吴洪宝等,2005)和独立样本检验两种方式来评估模型的预测技巧。在独立样本预测检验中,为了保证模型预测的独立性,增强模型的实际业务应用能力,将 1982—2009 期间作为训练阶段,确定预测因子,然后对 2010—2014 年作独立预测,检验模型的实际预测能力(表 2)。采用的评分办法有气候趋势预测评分 P_s ,距平符号一致率 Sc 和距平相关系数 Acc ,三种预测评定方法的定义见文献(陈桂英等,1998)。

从四种方案的交叉检验和回报验证结果可以看出(图 3 和图 4),四种方法的交叉检验都具有较好的评分结果,但是其独立样本检验的结果存在较大的差异。在方案 1 中,只使用前冬的海温资料对最近 5 年的独立样本回报检验中效果较差,其最近 5 年的回报的 Acc 均为负值。而采用海温变温方案即前期冬春季 3 月份的海温减去 1 月的海温的分区预测相对方案 1(前冬海温的年际变化)其预测效果有一定的提升,但是其回报检验的 Acc 仍然不稳定,如 2010 年 Acc 为明显负值。在方案 3 中增加了前冬 SLP 场提取的因子后,建模的回报验证结果的评分明显提升,交叉验证的结果变化不大,且独立样本检

验的结果较为稳定,距平符号一致率为 58%, Acc 近 5 年均值达到了 0.19,且近 5 年的回报检验结果均为正值,表明其预测结果较为稳定,无评分不理想的年份,预测结论有一定的参考价值。在方案 4 中,因子场不变,只是针对每个分区内的站点重新建模预

测,评分的值较分区预测有所降低,分区预测因子提取后对分区内单站的预测在趋势评分中要差于分区预测,这可能是降水趋势一致的区域能过滤单站的局地信息,区域代表性更为显著所造成的。

表 1 四种方案的因子选择和预测对象组合

Table 1 Four different schemes of factor selection and combination

方案序号	所采用的因子场	回报对象
方案 1	前期 1—2 月平均海温	分区降水距平百分率均值
方案 2	前期 3 月海温减去 1 月海温	分区降水距平百分率均值
方案 3	前期 3 月减去 1 月海温+前期 12—1 月平均 SLP	分区降水距平百分率均值
方案 4	前期 3 月减去 1 月海温+前期 12—1 月平均 SLP	单站降水距平百分率均值

表 2 四种方案的交叉检验和独立预测检验

Table 2 Cross-validation and independent forecast verification for four schemes

方案序号	1982—2009 年交叉检验			2010—2014 年独立预测检验		
	P_s	符号一致率	Acc	P_s	符号一致率	Acc
方案 1	80.4	0.61	0.29	65.5	0.44	-0.12
方案 2	81.4	0.63	0.34	76.5	0.56	0.09
方案 3	81.4	0.63	0.35	77.1	0.58	0.19
方案 4	76.3	0.59	0.24	72.1	0.54	0.12

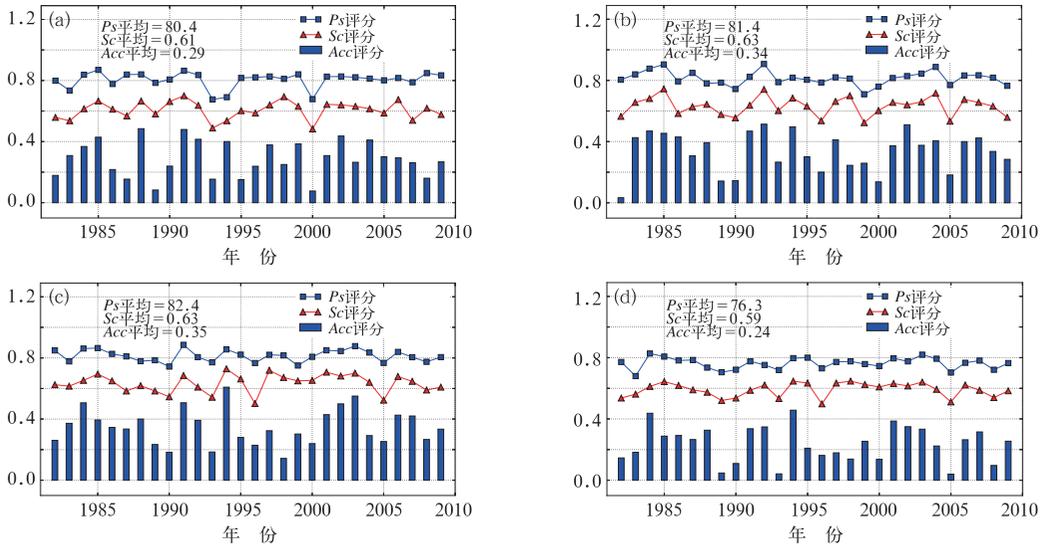


图 3 四种方案的交叉验证结果 (a~d): 预测方案 1~4

Fig. 3 Cross-validation (a-d) for four schemes 1-4

以江南东部区第 14 分区(图 5)为例,该区域位于我国东南部,这与我国南方夏季东南地区显著低频降水正中心区域接近(张玉洁等,2014)。而相关的研究表明前期海温的变化对该区域的降水趋势有一定的指示意义(黄丽娜等,2013)。针对该区域利用前期冬季海温年际异常变化预测降水,效果

不理想(图 6a);而利用海温冬春季演变的特征来提取信号,对该区域的夏季降水预测改进较为明显,5 年的回报检验显示,降水趋势正确的年份为 4 年(图 6b);加入冬季海平面气压场的信息后 5 年回报检验趋势正确的同样为 4 年,与利用海温变率作为因子结果一致(图 6c)。

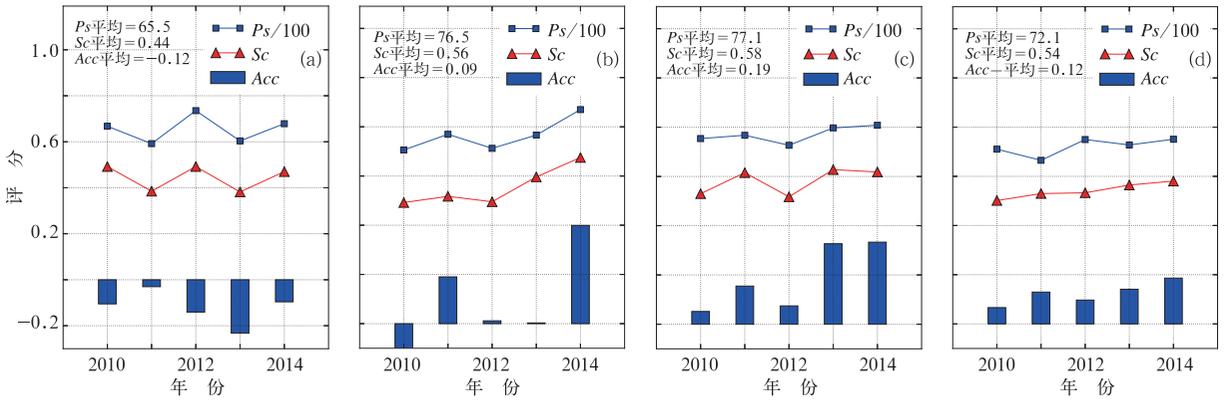


图 4 四种方案的独立样本检验结果 (a~d): 预测方案 1~4

Fig. 4 Retrospective forecast verification (a~d) for four schemes 1~4

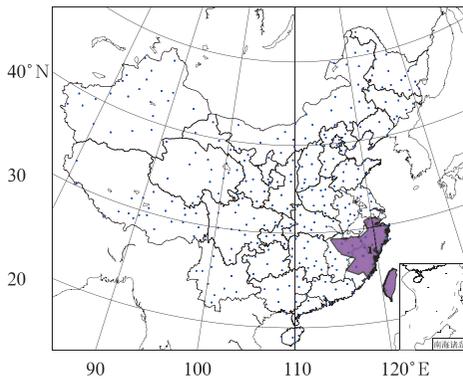


图 5 第 14 分区(江南东部区)站点分布

Fig. 5 Station distribution in Cluster 14

2014 年夏季我国东部降水主雨带位于江南地区, 江汉、黄淮地区降水明显偏少, 夏季气候呈南涝北旱的分布特征(图 7e)。方案 2 的 Acc 得分最高

为 0.40, 方案 3 和方案 4 分别为 0.33、0.19, 而方案 1 为一 0.10。与实况相比, 方案 1 对江南地区主雨带预测存在明显偏差(图 7a), 利用海温冬春差异的年际变化(方案 2), 可以预测出江南多雨的特征, 整个中国区域夏季的雨型基本正确, 预测的异常少雨中心在华北北部, 这与实况观测较为接近(图 7b)。增加前期海平面气压场作为预测因子, 长江以南的降水距平百分率的分布特征保持不变, 对黄河流域的预测有所改进, 河套西部区域的多雨趋势与实况接近(图 7)。分析聚类区域预测的结果, 可以得出分区预测模型对降水的异常预测能力不足, 这是由于分区降水的均值削弱了降水异常的历史信号造成的, 可以利用分区聚类提取的预测因子, 对单站的降水预测(方案 4)可以改进对降水异常趋势的预测能力(图 7d)。

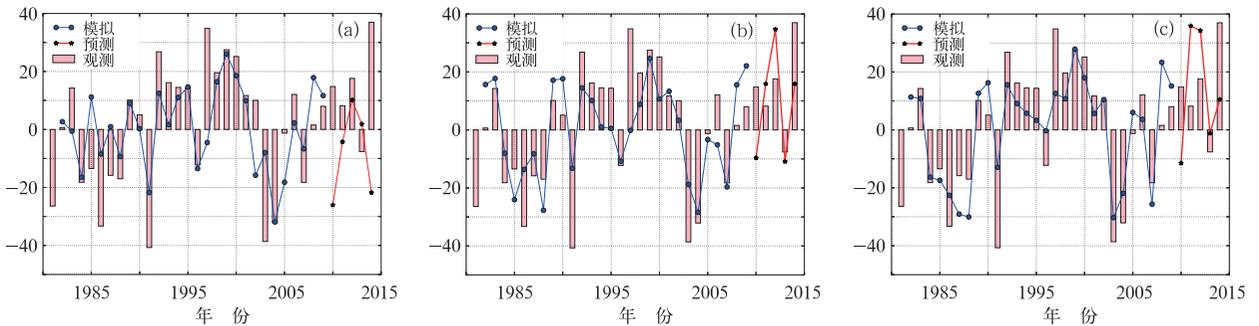


图 6 江南东部区域不同方案夏季降水回报检验

(a) 方案 1, (b) 方案 2, (c) 方案 3

Fig. 6 Reforecast verification for different schemes in eastern Jiangnan Region of China

(a) Scheme 1, (b) Scheme 2, (c) Scheme 3

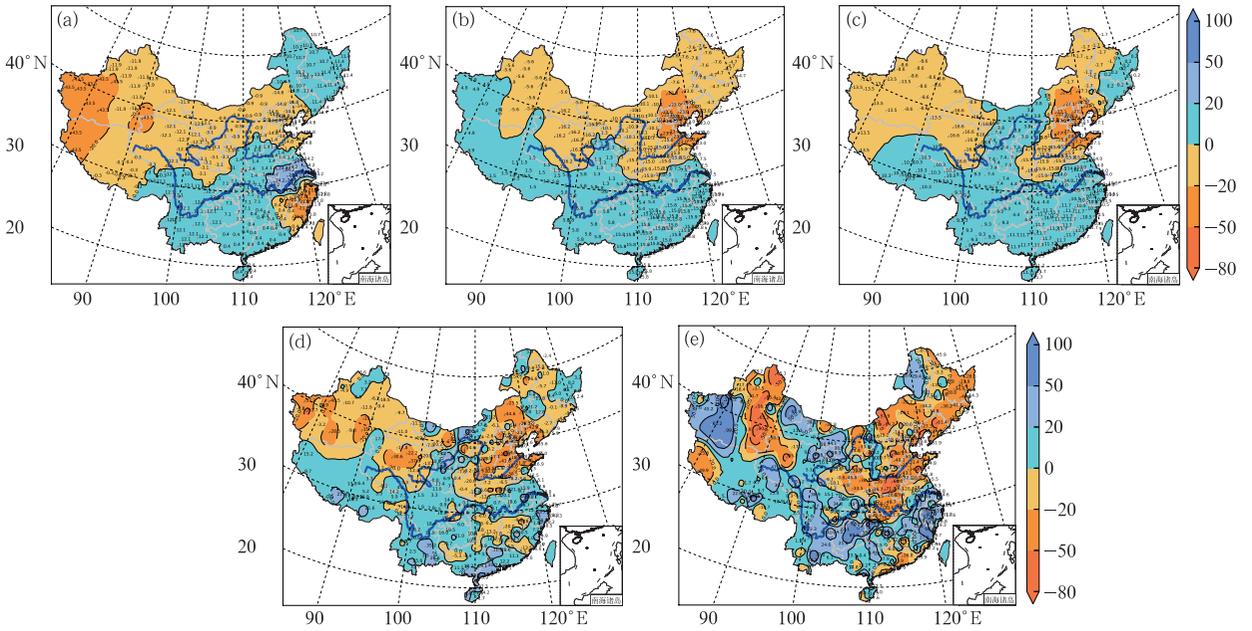


图 7 2014 年夏季降水回报与实况
(a~d) 预测方案 1~4, (e) 观测

Fig. 7 Reforecast verification for different schemes (a~d) and observation (e) for summer precipitation in China

5 结论与讨论

本文设计了一种基于客观聚类预测对象以及相关因子提取的统计预测方法,该方法基于近邻传播聚类综合考虑了相关系数、距平符号一致率和空间球面大圆距离对中国夏季降水进行分区,以不同分区的降水作为预测对象,分别选取前期冬春季的海温和海平面气压场为预测因子,利用图像标签算法提取因子场中高相关区域的预测因子信息,采用最小二乘回归法对中国不同分区的夏季降水趋势作预测,取得了较好的回报效果。对该预测模型的检验分析得到如下结论:

(1) 在全国的夏季降水趋势预测中,针对单个分区提取的预测因子信息,分别对分区预测和分区内不同单个站点回归预测,显示基于分区的降水趋势预测的回报验证得分要优于针对单站的预测得分,这可能是区域信息能过滤单站的局地不一致的干扰信息,区域代表性更为显著造成的,这说明利用一致性更强的区域气候特征作为预测的对象,可能具有一定预测技巧上的优势,针对分区寻找的前期的影响信号,对分区的降水趋势预测指示意义更强,但是对于其影响的物理机制和原因还有待于进一步深入探讨与分析。

(2) 几种不同的海气因子组合方案的预测分析结果表明,仅用前期冬春的海温的年际变率对中国区域的夏季降水预测能力不足,近 5 年的回报效果不理想。但是利用冬春季海温的演变特征并结合海平面气压的年际变化可以提高模型的预测技巧。

(3) 使用前期冬春海温演变的年际变化和前期海平面气压场组合的预测方案,能较好地预测出中国中东部 2014 年夏季降水南多北少的分布特征,预测的异常少雨中心在华北北部,整个中国区域夏季的降水距平百分率的预测值与实况观测分布较为接近。对应本文中分区预测的方案三的 2014 年夏季降水回报评分, P_s 得分为 80.8, 距平符号一致率为 62%, 距平相关系数为 0.33, 这样的预测效果对实际的业务应用具有一定的参考价值。

本文基于不同预测方案下开展中国夏季降水分区预测研究,虽然得到了一些有意义的结果,但需要指出本文的研究仍是初步的,其他聚类方法和分区方案在预测中的应用效果有待进一步验证。不同预测方案下,冬春季海温的变温特征以及大气环流因子对预测效果的改进,其原因有待于进一步的揭示。此外,本文的预测模型中仅考虑了前期海洋大气因子的可能影响,如何更好地结合海气耦合模式的信息采用本方法进一步提升该预测模型的能力也有待于进一步深入研究。

致谢: 本文得到了中国气象局短期气候预测创新团队的大力支持, 特此致谢!

参考文献

- 蔡榕硕, 谭红建, 黄荣辉. 2012. 中国东部夏季降水年际变化与东中国海及邻近海域海温异常的关系. *大气科学*, 36(1): 35-46.
- 陈桂英, 赵振国. 1998. 短期气候预测业务评估方法和业务初估. *应用气象学报*, 9(2): 178-185.
- 陈丽娟, 高辉, 龚振淞, 等. 2013. 2012 年汛期气候预测的先兆信号和应用. *气象*, 39(9): 1103-1110.
- 陈兴芳, 赵振国. 2000. 中国汛期降水预测研究及应用. 北京: 气象出版社. 9.
- 丁裕国, 张耀存, 刘吉峰. 2007. 一种新的气候分型区划方法. *大气科学*, 31(1): 129-136.
- 杜良敏, 张培群, 周月华, 等. 2011. 基于非负矩阵和典型相关的场定量预测方法研究. *高原气象*, 30(1): 103-108.
- 范丽军, 符淙斌, 陈德亮. 2007. 统计降尺度法对华北地区未来区域气温变化情景的预估. *大气科学*, 31(5): 887-897.
- 顾伟宗, 陈丽娟, 李维京, 等. 2012. 降尺度方法在中国不同区域夏季降水预测中的应用. *气象学报*, 70(2): 202-212.
- 顾伟宗, 陈丽娟, 张培群, 等. 2009. 基于月动力延伸预报最优信息的中国降水降尺度预测模型. *气象学报*, 67(2): 280-287.
- 郭玲, 何金海, 祝从文. 2012. 影响长江中下游夏季降水的前期潜在预报因子评估. *大气科学*, 36(2): 337-349.
- 郭艳君, 丁一汇. 2014. 1958~2005 年中国高空大气比湿变化. *大气科学*, 38(1): 1-12.
- 韩微, 翟盘茂. 2015. 三种聚类分析方法在中国温度区划分中的应用研究. *气候与环境研究*, 20(1): 111-118.
- 黄丽娜, 高建芸, 孙健, 等. 2013. 西北太平洋台风累积动能气候异常特征分析. *气象*, 39(8): 995-1003.
- 江志红, 丁裕国. 1994. 近 40 年我国降水量年际变化的区域性特征. *南京气象学院学报*, 17(1): 73-78.
- 金荣花, 田伟红, 矫梅燕. 2007. 基于 Ward 聚类法的中国业务集合预报系统的产品开发. *气象*, 33(12): 9-15.
- 柯宗建, 王永光, 龚振淞. 2014. 2013 年汛期气候预测的先兆信号及其应用. *气象*, 40(4): 502-509.
- 柯宗建, 张培群, 董文杰, 等. 2009. 最优子集回归方法在季节气候预测中的应用. *大气科学*, 33(5): 994-1002.
- 李维京, 丑纪范. 1999. 中国月平均降水场的时空相关特征. *高原气象*, 9(3): 284-292.
- 刘绿柳, 孙林海, 廖要明, 等. 2011. 基于 DERF 的 SD 方法预测月降水和极端降水日数. *应用气象学报*, 22(1): 77-85.
- 刘伟东, 尤焕琴, 任国玉, 等. 2014. 北京地区自动站降水特征的聚类分析. *气象*, 40(7): 844-851.
- 刘杨, 韦志刚, 李振朝, 等. 2012. 中国北方地区降水变化的分区研究. *高原气象*, 31(3): 638-645.
- 马京津, 李书严, 王冀. 2012. 北京市强降雨分区及重现期研究. *气象*, 38(5): 569-576.
- 毛炜峰, 南庆红, 史红政. 2008. 新疆气候变化特征及气候分区方法研究. *气象*, 34(10): 67-73.
- 吴洪宝, 吴蕾. 2005. 气候变率诊断和预测方法. 北京: 气象出版社, 144.
- 王会军, 范可, 朗咸梅, 等. 2012. 我国短期气候预测的新理论、新方法和新技术. 北京: 气象出版社, 120-121, 125-139.
- 王会军, 张颖, 郎咸梅. 2010. 论短期气候预测的对象问题. *气候与环境研究*, 15(3): 225-228.
- 王开军, 张军英, 李丹, 等. 2007. 自适应仿射传播聚类. *自动化学报*, 33(12): 1242-1246.
- 肖科丽, 赵国令, 方建刚, 等. 2015. 陕西汛期降水年际增量预测新技术研究. *气象*, 41(3): 328-335.
- 肖宇, 于剑. 2008. 基于近邻传播算法的半监督聚类. *软件学报*, 19(11): 2803-2813.
- 徐国昌, 姚辉. 1989. 我国降水量的模糊聚类分区. *气象*, 15(9): 13-17.
- 徐志清, 范可. 2012. 冬季和春季印度洋海温异常年际变率模态对中国东部夏季降水的可能影响过程. *大气科学*, 36(5): 879-888.
- 袁媛, 杨辉, 李崇银. 2012. 不同分布型厄尔尼诺事件及对中国次年夏季降水的可能影响. *气象学报*, 70(3): 467-478.
- 赵超, 舒红, 朱欣焰, 等. 2008. 气象数据概化中的最佳聚类数研究. *华中师范大学学报(自然科学版)*, 42(3): 490-499.
- 赵汉光, 张先恭. 1993. 我国东部夏季雨带的气候分类及其环流特征. *气象*, 9, 3-8.
- 张家诚. 1991. 中国气候总论. 北京: 气象出版社, 257-274.
- 张小曳. 2014. 中国不同区域大气气溶胶化学成分浓度、组成与来源特征. *气象学报*, 72(6): 1108-1117.
- 张玉洁, 刘寿东, 任宏利, 等. 2014. 中国南方夏季低频雨型特征及其年代际变化研究. *气象学报*, 72(6): 1205-1217.
- Chen Lijuan, Chen Deliang, Wang Huijun, et al. 2009. Regionalization of Precipitation Regimes in China. *Atmos Ocea Sci Lett*, 2(5): 301-307.
- Frey B J, Dueck D. 2007. Clustering by passing messages between data points. *Science*, 315(5814): 972-976.
- Kalnay E, Kanamitsu M, Kistler R, et al. 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bull Amer Meteor Soc*, 77: 437-471.
- Reynolds R, Rayner N A, Smith T M, et al. 2002. An improved in situ and satellite SST analysis for climate. *J Climate*, 15: 1609-1625.