

基于多模式的巢湖降水集成预报检验及应用

刘汉武 王文本 范裕祥 金社军 王玉红

(安徽省巢湖气象局, 合肥 238000)

摘要: 利用2014年2月—2015年9月中尺度模式预报(INCA、WRF)、全球模式(ECMWF、JMA)数据,结合巢湖湖泊周边区域站降水实况,应用自适应最小二乘集成法(简称统计集成)和加权平均集成法(简称加权集成)开展多模式集成预报试验,并对统计集成和加权集成效果进行检验与分析。研究表明:1)2种集成预报比单一模式预报的误差明显降低;2)在6~24 h时效内统计集成误差比加权集成误差大,在48~72 h时效内低于加权集成;3)在小雨级别,统计集成TS评分最高,中雨以上级别,统计集成TS评分低于加权集成,加权集成正确率始终高于统计集成。

关键词: 多模式, 加权集成, 统计集成, 降水, 巢湖

DOI: 10.3969/j.issn.2095-1973.2017.06.001

Preliminary Study of Precipitation Ensemble Prediction Based on Multimodel in Chaohu Lake

Liu Hanwu, Wang Wenben, Fan Yuxiang, Jin Shejun, Wang Yuhong

(Chaohu Lake Meteorological Service, Hefei 238000)

Abstract: Based on the data from INCA, WRF(mesoscale model), ECMWF, JMA (Global model) and observed data around Chaohu Lake, multimodel test equations are established to access the following two methods: the Adaptive least square statistical ensemble method and the weighted average method. Results show: 1) The error of forecasts from the 2 integrated forecasts is significantly small than that from the single model. 2) The error resulted by the Adaptive least square method is larger than that by the weighted average method in 6-24 h term, but is small in 48-72 h term. 3) For the light rain grade the, TS score of the Adaptive least square method was the highest. For above moderate rain grads, the TS score of the weighted average method is higher than other one. In addition, the correct rate of the weighted average method is always higher than that of the Adaptive least square method.

Keywords: multi-model, weighted average, adaptive least square, rainfall, Chaohu Lake

0 引言

目前短期数值预报实践中,模式的物理表达已经相当不错,一般要素的误差有相当大的部分源于初值的误差而不是模式本身的误差,想从单一模式着手来提高预报的准确率已经非常困难,使用两个或两个以上模式进行多模式集成试验,是近年比较流行的做法,它既考虑了初始误差的影响,又使预报结果更加稳定和可靠^[1]。目前比较流行的集成方法有加权平均集成法、线性回归集成预报法、多元回归集成预报法、最优权重集成预报法和人工智能神经网络集成法等。马清等采用算术平均和多元回归两种方法对2 m温度预报进行集成,结果表明两种集成方法的温度预报结果都优于单一模式预报^[2]。江滢等使用神经网络

集成方法对风速进行集成订正试验,结果发现集成订正效果优于任何一种单一的统计订正,集成订正都能较好地消除系统误差,具有更为明显的平均误差小优势^[3]。刘静等利用ECMWF、JMA、WRF和MM5做淮河流域面雨量多模式预报效果比较,结果表明随着雨量等级增大,各模式TS评分均明显下降,ECMWF评分最高,JMA次之^[4]。冯汉中基于TS评分权重系数进行集成预报试验比,结果显示以多模式TS评分为权重系数的集成预报优于逐步回归和支持向量(SVM)回归方法,多个模式TS评分为权重系数集成的降水预报效果相对于单一模式有明显提高^[5]。农孟松等运用人工神经网络与主分量分析(PCA)相结合的方法,对同一降水预报量的各种数值预报产品进行集成预报研究,结果表明集成预报对几种降水数值预报产品有较好的集成优化效果,并优于各子预报产品^[6]。彭九慧等采用多数表决集成法、评分权重集成法、多元回归集成法进行集成预报试验。结果表明:在单一预报中TS值均较高,多数表决集中集成法可大幅度提高预

收稿日期:2016年9月30日;修回日期:2017年5月8日
第一作者:刘汉武(1970—),Email:liuhanwu226@sohu.com
资助信息:安徽省预报员专项科研项目(KY201504);
欧洲细网格预报在巢湖流域内精细化订正与应用

报准确率, 如果单一预报准确率相差较大时, 采用评分权重集成法提高准确率, 在样本足够的条件下, 多元回归集成法最好^[7]。程鹏等采用并集集成方法进行了降水集成预报试验, 结果表明并集集合的预报能力比单个成员明显增强, 集成预报效果总体上比任何单一成员的预报效果好^[8]。

巢湖湖泊是我国第五大淡水湖, 水域面积近800 km², 水体的气象要素和陆地有明显的差别, 进行巢湖降水集成预报的研究, 对保障巢湖水上交通、渔业、旅游和水利调度等活动有着重要的意义。本文应用统计集成和加权集成进行集成预报试验, 探讨合适巢湖降水预报集成的方法, 以期提高预报准确率。

1 资料和方法

目前巢湖湖区及沿湖共29个自动雨量站点(蓝色为2014年底前投入使用站点), 分布如图1。本文使用的区域自动站观测数据是经安徽省气象局质量控制后的自动站数据, 按照站点投入使用时间顺序, 逐步加入试验。用平均法、反距离加权插值法、泰森多边形法计算出的面雨量对比, 差别不明显, 最后选择泰森多边形法计算出的面雨量参与集成试验。

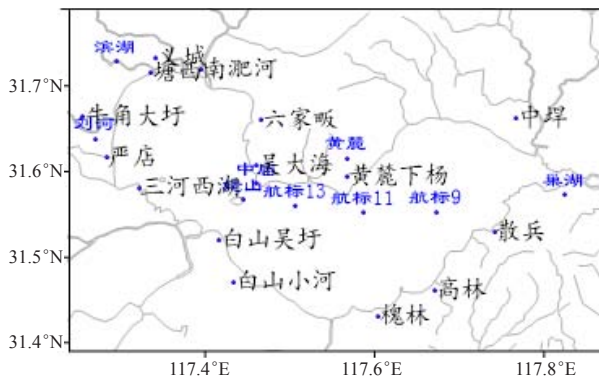


图1 巢湖周边自动气象站分布图

Fig. 1 Locations of automatic weather stations around Chaohu Lake

1.1 资料

选取2014年2月—2015年9月环巢湖站点泰森多边形面雨量、ECMWF、JMA、安徽WRF和INCA中尺度预报资料。对巢湖湖面及其沿湖站点用泰森多边形法算出巢湖降水实况数据; ECMWF在2015年1月12日前分辨率为0.25°×0.25°, 对巢湖区域6个点求平均, 2015年1月12日以后的分辨率为0.125°×0.125°, 对巢湖区域的9个点求平均, 得到ECMWF预报值; JMA分辨率为0.5°×0.5°, 挑选巢湖区域1个点作为日本预报值; WRF分辨率为0.025°×0.025°, 对巢湖湖区内102个数据求平均值作为WRF预报值; INCA分辨率为

0.01°×0.01°, 对巢湖区域843个点数据求平均, 得出INCA的降水数据。

1.2 集成方法

1.2.1 统计集成法

用预报时间段的前 m 天ECMWF、JAM、WRF、INCA模式预报数据及其对应巢湖面雨量数据, 按时间顺序构建训练样本, 来进行统计集成。不同样本数集成预报误差也不同, 只要挑选出集成预报误差最小的样本数, 就是最佳统计集成预报, 可表达为

$$\sum_{j=1}^n X_{ij}\beta_j = y_i \quad (i=1, 2, 3, \dots, m) \quad (1)$$

1.2.2 加权集成法

加权集成法原理比较简单, 但降水预报特点仅限于有和无, 采用该方法可以有效突出各模式降水预报集中特点, 提高预报准确率, 但很难报准极端的大级别降水天气。在推算集成预报方程时, 尽量用较多样本, 计算出集成方程权重系数才具有代表性。本文中使用的环巢湖实况资料建站年限(2014年2月)较短, 本次试验采用动态变更训练样本, 即每年变更一次训练样本, 确定次年预报方程, 在训练样本足够的情况下, 方程趋于稳定, 就可以固定方程。在方程确定后, 只要有各模式预报结果即可进行集成预报。加权集成公式为:

$$y(t) = \sum_{i=1}^k a(t)_i x_i(t) \quad (2)$$

式中, $a(t)_i$ 为权重系数。确立了权重系数, 就可以利用每天的各家模式预测结果进行多模式降水集成预测。

1.3 检验方法

依据中国气象局《中短期天气预报质量检验办法》, 采用均方根误差检验、正确率(PC)检验和TS评分和空报率、漏报率对各单一模式和集成预报进行效果检验, 降水量分级检验分为小雨、中雨、大雨、暴雨以上几个等级, 检验时效为6、24、48和72 h。

$$\text{正确率: } PC = \frac{NA + ND}{NA + NB + NC + ND} \times 100\% \quad (3)$$

式中, NA 为有降水预报正确站(次)数, NB 为空报站(次)数、 NC 为漏报站(次)数, ND 为无降水预报正确的站(次)数。

$$\text{TS评分: } TS_k = \frac{NA_k}{NA_k + NB_k + NC_k} \times 100\% \quad (4)$$

$$\text{漏报率: } PO_k = \frac{NC_k}{NA_k + NC_k} \times 100\% \quad (5)$$

$$\text{空报率: } FAR_k = \frac{NB_k}{NA_k + NB_k} \times 100\% \quad (6)$$

式中, NA_k 为等级降水正确(即预报等级与实况等级

相同)的天数; NB_k 为空报站天数(即预报等级大于实况等级,记录为等级空报); NC_k 为漏报站天数(即预报等级小雨实况等级,记录为实况等级漏报)。

2 集成与检验

2.1 预报集成

用统计集成来试验巢湖降水集成预报,不同训练样本长度集成预报误差也不同,以24 h为例,随着样本数的增加,统计集成法预报的均方根误差(图2)先下降,在第137个样本数时达到最低值5.992 mm,后呈缓慢上升趋势,故137样本即为最小二乘法巢湖降水集成最佳样本。

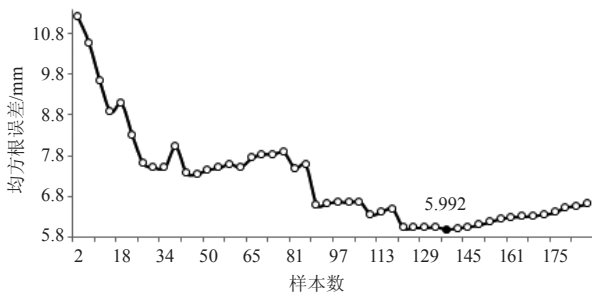


图2 24 h统计集成训练样本和均方根关系

Fig. 2 Relationship of the root-mean-square (RMS) error to the training samples in 24-hour forecasts

同理得到其他时效的最佳样本天数,统计结果如表1。

表1 最佳统计集成预报的样本天数和均方根误差(单位: mm)
Table 1 The best training sample and root mean square error of the least squares integrated prediction (unit: mm)

时效/h	6	24	48	72
最佳训练样本天数/d	110	137	133	133
均方根误差/mm	3.222	5.992	5.888	8.858

加权集成法,6 h采用4家模式参与降水集成试验,24 h以上采用ECMWF、JMA和WRF参与集成。以24 h为例,用2014年2月到2015年9月试验数据。当集成预报的均方根误差均低于ECMWF均方根误差时,把ECMWF模式在集成预报中权重从0.42变化到1,剩下部分权重平分给JMA和WRF,结果显示,ECMWF权重为0.72时,均方根误差达到最小,为5.932 mm(图3a),确定ECMWF权重为0.72。固定ECMWF的权重,然后把JMA的权重从0变化到0.28,WRF权重则从0.28变化到0(图3b),发现JMA和WRF同时为0.14时,集成预报均方差误差刚好达到最小值5.932 mm。

因此,得到24 h最佳集成预报方程:

$$y = 0.72 \times ECMWF + 0.14 \times JMA + 0.14 \times WRF; \quad (7)$$

同理,得到6 h最佳集成预报方程为:

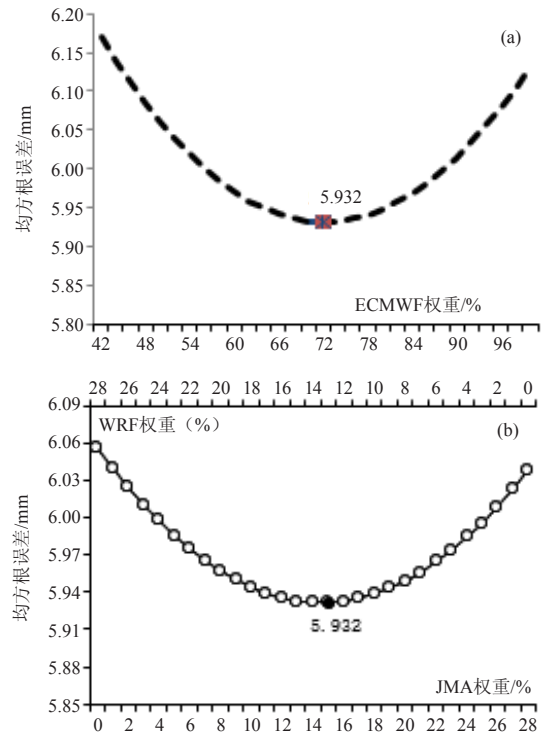


图3 加权集成法均方根误差和权重系数变化
(a) ECMWF; (b) JMA和WRF

Fig. 3 Relationship of the RMS error to the weight-coefficients for:
(a) ECMWF; (b) JMA and WRF

$$y = 0.41 \times ECMWF + 0.18 \times JMA + 0.14 \times WRF + 0.27 \times INCA; \quad (8)$$

48 h最佳集成预报方程为:

$$y = 0.52 \times ECMWF + 0.21 \times JMA + 0.27 \times WRF; \quad (9)$$

72 h最佳集成预报方程为:

$$y = 0.4 \times ECMWF + 0.34 \times JMA + 0.26 \times WRF. \quad (10)$$

由于各家模式都含有权重,实际工作中会出现个别模式预报降水,其他模式均没报降水。为进一步提高集成预报的准确率,采用多数表决议法干预集成预报结果,即当4家模式预报参与试验时,无降水站点 ≥ 3 ,认为集成预报结果无降水,如此调整集成预报结果,可以有效提高预报的准确率。用2014年2月—2015年9月资料,对各单一模式和24 h最佳集成预报结果进行全年和分季节预报,并结合多数表决议法,各单一模式和集成预报的均方根误差、正确率、空漏报率统计如表2,结果发现加权集成法的均方根误差均小于各单一模式,正确率高于大部分单一模式,漏报率明显降低,空报率略有提高。

2.2 均方根误差检验

从2014年2月—2015年9月组合ECMWF、JMA、WRF和INCA四种模式对巢湖湖区(6~72 h)降水预

表2 各单一模式和加权集成24 h均方根误差、正确率、空漏报率

Table 2 The root-mean-square error, the correct rates, vacancy rate, and false negative rate resulted by each single mode and Weighted average integration in 24 h-term

时间	项目	ECMWF	JMA	WRF	加权集成
春	空报率/%	15.3	11.8	12.2	12.1
	漏报率/%	20.6	28.6	31.7	19.0
	正确率/%	77.6	75.5	73.5	80.6
	均方根误差/mm	5.3	7.6	7.1	5.2
夏	空报率/%	18.9	19.0	18.0	20.0
	漏报率/%	8.0	12.5	26.7	7.1
	正确率/%	80.4	78.0	71.2	79.8
	均方根误差/mm	8.5	10.7	11.8	8.2
秋	空报率/%	23.9	22.6	20.5	25.0
	漏报率/%	24.0	10.9	32.6	15.2
	正确率/%	77.6	82.6	76.5	79.6
	均方根误差/mm	2.7	4.4	4.4	2.4
冬	空报率/%	0	29.6	0	0
	漏报率/%	42.3	26.9	46.2	30.8
	正确百分率/%	77.6	69.4	75.5	83.7
	均方根误差/mm	1.2	2.4	1.6	1.0
年	空报率/%	17.8	19.4	15.8	17.8
	漏报率/%	17.8	17.8	31.2	14.2
	正确率/%	78.7	77.5	73.6	80.4
	均方根误差/mm	6.1	8.1	8.6	5.7

报均方根误差检验结果（图4）可以看出：随着预报时效的延长，各模式预报和集成预报的均方根误差都呈上升趋势，6~72 h不同时效，统计集成和加权集成预报均方根误差均略低于单一模式预报，ECMWF均方根误差次之。6 h加权集成法预报误差最小，48、72 h统计集成预报均方根误差最小。

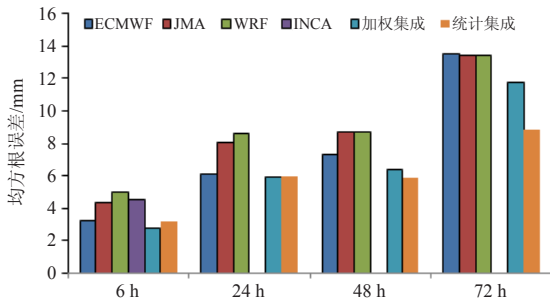


图4 ECMWF、JMA、WRF、INCA和集成预报均方根误差
Fig. 4 Root-mean-square error from the ECMWF, JMA, WRF, INCA and integrated forecast (unit: mm)

2.3 正确率 (pc) 检验

从2014年2月—2015年9月组合ECMWF、JMA、WRF和INCA四种模式对巢湖湖区（6~72 h）降水预报正确率检验结果（图5）可以看出：随着预报时效的延长，各模式预报和集成预报的正确率都呈下降趋

势。加权集成正确率始终高于统计集成正确率，加权集成24 h（80.4%）和72 h（68.9%）略高于单一模式预报，ECMWF 48 h（76.4%）略高于加权集成为最佳，INCA 6 h以88.6%的正确率明显大于其他模式，加权集成排在ECMWF之后位列第三。统计集成正确率一直表现一般，排名第三左右。

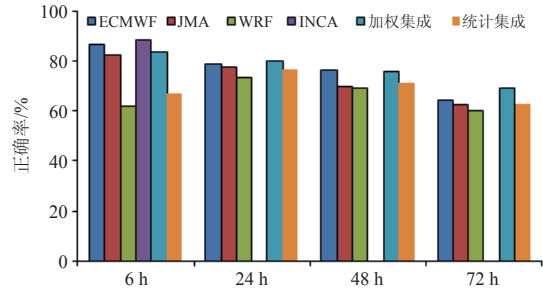


图5 ECMWF、JMA、WRF和INCA、加权集成、统计集成预报正确率

Fig. 5 The correct rate resulted from the ECMWF, JMA, WRF and INCA, weighted average and least square integrated prediction

2.4 TS 评分和空报率、漏报率检验

分级检验可以看出：1）随着降水等级的增大，24 h以后随着时效的延长，各家模式TS评分逐渐下降（图6a）。在TS评分中，小雨级别时，24 h后所有统计集成预报高于其他所有预报；中雨级别时，统计集成评分掉落很快，加权集成预报评分超过ECMWF，效果最好；大雨级别，ECMWF效果较好；暴雨级别时，WRF最好。72 h大雨和暴雨级别时，各模式预报效果都较差。6 h，集成预报最佳，INCA其次，WRF效果最差。TS评分总体评价，小雨级别时，统计集成预报最好；其他级别时ECMWF最好，加权集成预报其次。2）各模式和集成预报的空报率和漏报率都随着预报时效的延长和降水等级的增加逐渐增大。在小雨和中雨阶段集成预报漏报率较小，空报率偏大。24 h各模式小雨和中雨空漏报率低于45%，大雨和暴雨空报率和漏报率均超过60%，72 h的大雨和暴雨预报基本不准确。3）统计集成6 h中雨空报率和6 h中雨以上漏报率出现偏高，经统计发现，使用110天最佳样本长度预报，实况只出现5次中雨，均漏报。因此，在6 h中雨以上，如果大降水次数较少，不适合使用统计集成方法。

3 应用效果检验

利用最佳加权集成预报方程尝试对2016年1—6月巢湖面雨量进行预报试验，统计分析结果见表3。

通过实际检验，可以看出：随着时间延长，集成预报同样正确率逐渐降低，平均绝对误差逐渐增大。

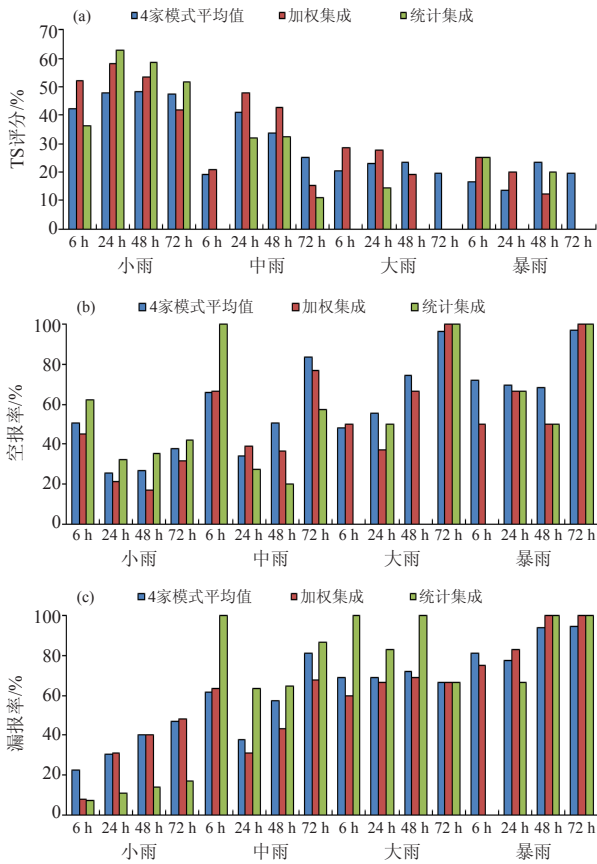


图6 2014年2月—2015年9月 ECMWF、JMA、WRF、INCA和集成预报TS评分(a)、空报率(b)、漏报率(c)
Fig. 6 TS score (a), vacancy rate (b), and false negative rate (c) by the ECMWF, JMA, WRF and INCA and integrated forecast for February 2014–September 2015

TS评分随着时效的延长和雨量级别增大, 评分逐渐降低。除6 h外, TS评分加权集成小雨评分均高于其他数值预报模式, 暴雨评分出现明显的上升, 集成预报在实际工作中有一定的实用价值。

4 结论

通过对ECMWF、JMA、WRF和INCA尝试集成预报检验, 得出以下几个结论:

1) 统计集成预报, 随着样本数的增加, 集成预报的均方根误差呈现先下降后缓慢上升的趋势, 训练样本数在110~140之间时, 均方根误差达到最小。最佳统计集成预报在6~24 h时比加权集成误差大, 48~72 h时比加权集成预报误差小。正确率则一致低于加权集成。在TS评分中, 在小雨级别, 统计集成预报效果最好。随着预报时效的延长, 在中雨以上时, 统计集成法的TS评分不如其他模式。

表3 2016年1—6月各模式和加权集成法预报效果检验
Table 3 Effect evaluation of various models and weighted average method for January–June 2016

预报时效/h	项目	正确率/%	平均绝对误差/ mm	TS评分			
				小雨	中雨	大雨	暴雨
6	加权集成	0.62	6.27	0.32	0.25	0.33	/
	ECMWF	0.65	3.37	0.30	0.25	/	/
	JMA	0.66	3.50	0.33	0.57	0	/
	WRF	0.71	10.20	0.14	0	/	/
	INCA	0.68	3.65	0.29	0.20	0	/
24	加权集成	0.57	7.67	0.37	0.35	0.29	0.67
	ECMWF	0.58	8.30	0.36	0.30	0.33	0.50
	JMA	0.57	5.80	0.32	0.16	0.33	0.33
	WRF	0.6	13.10	0.31	0.50	0.16	0
48	加权集成	0.47	12.50	0.30	0.26	0.13	0.50
	ECMWF	0.46	18.30	0.25	0.15	0.14	0.25
	JMA	0.43	11.70	0.27	0.20	0.06	0.50
72	加权集成	0.68	11.60	0.34	0.25	0.16	/
	加权集成	0.39	19.10	0.27	0.06	0.05	0.20
	ECMWF	0.41	28.30	0.24	0	0.10	0
	JMA	0.39	21.50	0.26	0.04	0.07	0.20
	WRF	0.64	12.50	0.32	0	0	0

2) 加权集成最佳预报方程如公式(7)~(10)所示。由于ECMWF预报稳定性较好, 在加权集成中权重最大, 其中24 h权重最高达0.72。在6 h检验中, INCA表现出较强的优势, 不仅其权重超过JMA和WRF, 正确率也高于其他模式和集成预报, 空报率较低, 但漏报率较高。

3) 两种集成方法最佳集成预报均方根误差均比任何一个单模式预报的误差小, 24和72 h集成预报正确率最高。在加权集成方程中, 结合多数表决多模式可以有效提高预报准确率; 多模式集成预报降低漏报百分率, 同时也提高了空报百分率。

参考文献

- [1] 祁海霞. 多模式集成预报及概率预报研究. 南京: 南京信息工程大学, 2010: 1-50.
- [2] 马清, 龚建东, 李莉, 等. 超级集成预报的误差订正与集成研究. 气象, 2008, 34(3): 42-48.
- [3] 江滢, 宋丽莉, 程兴宏. 风电场风速集成订正方法的尝试性研究. 资源科学, 2013, 35(3): 673-680.
- [4] 刘静, 叶金印, 张晓红, 等. 淮河流域汛期面雨量多模式预报检验评估. 暴雨灾害, 2014, 33(1): 58-64.
- [5] 冯汉中, 肖红茹, 徐琳娜, 等. 基于数值预报模式降水产品的集成应用方法试验. 高原山地气象研究, 2009, 29(3): 39-43.
- [6] 农孟松, 黄海洪, 孙崇智, 等. 基于主分量神经网络的降水集成预报方法研究. 气象, 2011, 37(3): 352-355.
- [7] 彭九慧, 丁力, 杨庆红. 几种降水集成预报方法的对比分析. 气象科技, 2008, 36(5): 520-523.
- [8] 程鹏, 郑启锐, 张涛. 数值降水预报结果的并集集成方法及其试验研究. 暴雨灾害, 2007, 26(3): 256-260.