

# “SEEPS” 降水预报检验评分方法在我国降水预报中的应用试验

陈法敬 陈静

(国家气象中心, 北京 100081)

**摘要:** 介绍了国际上一种新的降水检验方法——概率空间中的稳定公平误差 (stable equitable error in probability space, SEEPS) 的原理、计算方法和误差特征, 并应用于我国定量降水预报检验进行评估试验。SEEPS方法在评分意义、降水分类、评分计算及评分应用等方面, 比传统检验评分更加灵活, 具有更清晰的实际意义。利用两个降水概率阈值, SEEPS方法将降水气候概率分布划分为“干”、“小雨”、“大雨”三类; 该方法基于降水概率计算误差矩阵, 根据站点分布密度计算区域平均评分权重系数。SEEPS在不同降水概率下具有不同的误差评分特征, 使其能够自动适应不同的降水气候。SEEPS不仅可以定量化给出降水预报能力的高低, 还可以通过分析不同观测和预报分类组合的误差评分, 给出造成评分高低的成因。利用2007年3月—2013年12月24h累积降水逐日观测资料, 对中央气象台预报员定量降水预报进行了SEEPS检验试验, 并与传统的检验评分进行比较。结果表明: 预报员定量降水预报的误差主要来源为两类——预报“小雨”对观测“大雨”的漏报和预报“小雨”对观测“干”类型的空报, 合计占到了总误差的近七成; 前者说明预报员降水预报量级较实际降水偏小, 后者说明预报员对“小雨”的大范围、高频率空报也可以导致总体预报效果的明显下降。SEEPS方法对降水预报能力的评估结论与传统检验评分总体相当, 但SEEPS检验指标更简单直接, 便于管理层和决策层面使用, 具有较好的推广应用价值。

**关键词:** 降水预报, 检验方法, SEEPS, 概率检验

**DOI:** 10.3969/j.issn.2095-1973.2015.05.001

## The Application Experiment of a New Score for Precipitation Verification Based on the SEEPS Principle

Chen Fajing, Chen Jing

(National Meteorological Centre, Beijing 100081)

**Abstract:** The principles, computation and error characteristics of a new method newly developed internationally for precipitation verification, i.e., SEEPS (Stable Equitable Error in Probability Space), are briefly introduced in this paper. With regard to score meaning, precipitation classification, score calculation and application, SEEPS is more flexible and has clearer practical meaning than traditional verification scores. The SEEPS method is applied in the verification and assessment experiment for quantitative precipitation forecasts in China region, and some issues encountered in the application are talked about. Climatic precipitation probability distribution is divided into three classifications, ‘dry’, ‘light’ and ‘heavy’ by two thresholds in SEEPS. Error matrices are determined by the precipitation probability. Area mean is weighed based on the station density. Possessing different error score characteristics under varying precipitation probabilities makes SEEPS automatically adapt to various kinds of precipitation climate. The ability of precipitation forecasts is quantified by the SEEPS values, and different elements of SEEPS can also be analyzed to find the reason why the SEEPS score is high or low. Daily observations of 24-hour accumulated precipitation from Mar. 2007 to Dec. 2013 are used in the SEEPS verification test for quantitative precipitation forecasts (QPF) of forecasters in the National Meteorological Center of CMA. Results show that the two main categories of forecasters’ QPF errors are misses of heavy precipitation with light precipitation forecasts, and false alarms of “dry” with light precipitation forecasts. These two kinds of errors account for about 70 percent of all errors. The former explains the order of forecast precipitation is lower than observations and the latter states that false alarms with large areas and high frequencies can also obviously deteriorate the whole forecast performance. Conclusions of SEEPS are approximately equivalent to those of traditional verification scores. But the SEEPS score is simpler and more suitable for use in management and decision making, and has more value for popularization and application.

**Keywords:** precipitation forecast, verification method, SEEPS, probabilistic verification

收稿日期: 2014年1月4日; 修回日期: 2014年6月24日  
通信作者: 陈静 (1966—), E-mail: chenj@cma.gov.cn  
资助信息: 中国气象局关键技术集成与应用项目 (CMAGJ2014Z03);  
国家自然科学基金项目 (41075035); 中国气象局新技术推广项目 (CMAGJ2012Z01); 科技部国家科技支撑计划 (2009BAC51B00); 国家重点基础研究发展计划 (2012CB417204); 公益性行业 (气象) 科研专项 (GYHY200906007)

## 0 引言

定量降水是气象预报中最受关注的气象预报要素, 如何评估定量降水预报准确率, 从理论到实践均处于不断的发展、改进之中。目前, 业务降水预报检验主要分为预报员主观降水检验和客观统计降水检验两类。主观检验主要以单站降水观测为基础, 对比分析降水预报的时间、落区与强度误差等, 并分析其成因<sup>[1]</sup>。这种降水检验方法直观, 但不便于总体上客观、定量地评价降水预报能力。客观方法主要基于二分类事件[事件发生或不发生, 即(0, 1)事件]的传统检验评分方法, 采用列联表给出统计评分。这类降水检验方法, 可以从多个角度、不同量级检验、评估降水预报性能<sup>[2-6]</sup>, 但同样不能单独给出代表降水预报整体性能的一个定量指标。

近年来, 国际上新发展了多种降水预报检验方法。Ebert<sup>[7]</sup>研究了模糊检验, 该方法不追求预报与观测在位置上的精确匹配, 利用一个滑动的某一尺度的窗口, 对落在该区域中的所有降水观测与预报进行统计, 优点是插值问题就不再存在, 允许预报存在一定的空间位置偏差。新兴起的空间检验方法也获得了广泛的研究与应用<sup>[8]</sup>, 如Davis等<sup>[9-11]</sup>研究的MODE方法, 通过获取观测与预报的降水落区形状, 计算两者空间位置、面积和轴向等要素的差异, 该方法是主观天气学检验的客观量化。此外, 国际上还新发展了一些与决策、经济价值相关联的预报检验方法<sup>[12]</sup>, 及针对极端降水等小概率事件的检验方法<sup>[13]</sup>。

一般说来, 业务检验评分倾向于应用简单、客观的指标, 而且是越简单越好。由于传统的降水检验评分需要划分降水等级, 不便于评估比较不同气候区域特征的降水预报性能, 2010年欧洲中期天气预报中心(ECMWF)的Rodwell等<sup>[14]</sup>研究了一种新的降水检验方法——概率空间中的稳定公平误差(stable equitable error in probability space, SEEPS), 以克服降水检验中分类较多及受气候特征影响大的问题。传统分量级统计检验中, 不同量级检验结果不一致时, 就难以确定总体检验结果的好坏。在不同气候区域, 同一量级的降水发生概率相差较为悬殊时, 传统的统计检验评分就不适合进行相互比较。SEEPS方法结果是一个综合评分, 其降水分类以降水概率为基础进行划分。相比传统检验评分, SEEPS方法更好地处理了上述两种情况。2012年, Haiden等<sup>[15]</sup>利用SEEP方法, 评估了ECMWF、英国气象局(UKMO), 美国环境预报中心(NCEP)、日本气象厅(JMA)和加拿大气象局(CMC)等五个国家和地区的数值

预报模式1~10d的降水预报性能, 通过与传统降水检验方法进行对比, 分析了不同模式降水预报能力的差异。在对降水预报的检验中, 与传统二分类降水检验方法相比, 基于降水气候概率的降水检验新方法SEEPS具有一些明显不同的特征。2011年世界气象组织(WMO)在第十六届大会上, 确定ECMWF作为确定性数值预报检验领导中心(Lead Centre for Deterministic Forecast Verification, LC-DNV), 制定了确定性数值预报检验的一套标准评分的制作与交换流程, 确保不同中心数值预报产品的检验信息的一致性。现在SEEPS降水检验方法已在LC-DNV中获得业务应用, 可对ECMWF、UKMO、NCEP和JMA等四个数值预报中心全球数值模式的24h累积降水预报分区检验, 检验结果通过ECMWF的网页(<http://apps.ecmwf.int/wmolcdnv/scores/surface.mean/tp>)查询显示。

在评分意义上, 传统二分类检验方法, 包括TS/ETS评分、预报偏差(Bias)、空报率、漏报率等, 是利用某一降水量级上降水分类列联表的四类观测—预报情况的站点数 $a$ 、 $b$ 、 $c$ 、 $d$ , 近似代表四类观测—预报情况面积, 构成评分指标, 以此来考察预报某一方面的性能<sup>[2]</sup>。首先, 这类评分指标依赖于站点的分布状况; 高站点密度的区域, 可能会主导整个评分。其次, 这类评分的计算形式, 致使其逐日评分的时间均值, 与利用整段时间内的检验样本计算获得的评分并不相等。因此, 不同的评分计算方式可能会得到两种不同的检验结论。SEEPS方法不同于此, 它是基于降水气候概率, 在概率空间中将降水分为三类, 并对降水分类 $3 \times 3$ 列联表中不同观测—预报情况给予不同的SEEPS误差评分, 计算概率空间中基于降水气候概率的预报误差。SEEPS评分的时间均值与该时间段内的SEEPS评分是一致的, 并且与区域平均计算的先后顺序无关。通过降低高站点密度区域的站点权重, 以避免高站点密度区域主导区域平均SEEPS评分。

在评分计算上, 传统二分类检验方法需要对降水量级分级(一般分为小雨、中雨、大雨、暴雨和特大暴雨五级), 计算降水分类 $2 \times 2$ 列联表中的四类观测—预报情况站点数, 由此计算各检验评分。SEEPS检验方法中, 也对降水进行了分类(“干”、“小雨”、“大雨”), 但依据是降水气候概率, 对降水分类 $3 \times 3$ 列联表中不同观测—预报分类给予不同的SEEPS误差评分, 由此获得一个可以代表降水预报整体性能的单—评分。

在评分应用上, 判断和评估一种降水预报的整体性能时, 若利用传统检验评分, 需要综合不同量级上

的不同检验评分，做出一个定性的、描述性的预报检验结论。经常遇到下面的两种情况，两种预报的同一传统检验评分（如TS评分）在不同降水量级上的大小次序是相反的；同一降水量级上，两类不同的传统检验评分（如空报率与漏报率）检验结果相反（如漏报率低、空报率高的预报）。利用SEEPS检验方法获得的是单一检验评分，可以直接定量地给出不同降水预报整体预报能力的差异，这对于一些依赖于预报性能判断、评价的决策是很方便的。同时，通过分析构成SEEPS误差评分及观测—预报列联表的九个元素，可以判断、理解观测—预报的九个不同分类对SEEPS误差的贡献，显示预报中存在的主要缺陷，为改进预报性能提供指导信息。

本文将SEEPS方法应用到中国区域降水预报检验中，简要介绍SEEPS方法的原理，对中央气象台预报员2007—2013年的定量降水预报进行了评估试验，并与传统检验评分进行对比分析，对SEEPS应用中存在的问题进行了讨论。

## 1 SEEPS评分原理及计算方案简介

### 1.1 基于概率分布特征的降水分类方法

将某一站点降水按其气候概率分布特征划分为“干”、“小雨”、“大雨”三类，定义其概率分别为 $P_1$ 、 $P_2$ 和 $P_3$ ，三类降水概率满足条件： $P_1+P_2+P_3=1$ ，设 $P_1$ 定义为24h累积降水 $\leq 0.2\text{mm}$ 的累积概率，“小雨”降水概率 $P_2$ 为“大雨”降水概率 $P_3$ 的两倍，即 $P_2=2P_3$ 。当“干”的气候概率 $P_1$ 确定后，小雨和大雨的概率值也就随之确定。按概率进行分类，可以减小观测降水样本的异常值影响，降低对降水观测样本的质量控制要求。

设降水观测与预报分别为 $v$ 和 $f$ ，对“干”、“小雨”和“大雨”三类降水事件，定义一个 $3 \times 3$ 列联表，如表1所示。该列联表为降水观测与预报的二维离散概率分布，代表SEEPS检验方法的降水类别及预报特征。若是“完美预报”，则列联表对角线元素之和为1： $P_{11}+P_{22}+P_{33}=1$ ，非对角线元素均为0；若是无技巧的随机预报，则降水观测的条件概率分布 $P(v|f)$ 与边缘分布 $P(v)$ 相同，与参数 $f$ 无关；即 $P(v|f=1)=P(v|f=2)=P(v|f=3)=P(v)$ 。

表1 SEEPS检验方法的降水分类预报与观测 $3 \times 3$ 列联表

		观测			
概率		$P_1$	$P_2$	$P_3$	
		$v$			
Cat		1	2	3	
预报	$f$	1	$P_{11}$	$P_{12}$	$P_{13}$
	2	$P_{21}$	$P_{22}$	$P_{23}$	
	3	$P_{31}$	$P_{32}$	$P_{33}$	

### 1.2 SEEPS方法原理简介

Gandin等<sup>[16]</sup>定义了一个高效的检验评分，如下：

$$S = \sum_{f,v} P_{fv} S_{fv} \quad (1)$$

式中， $P_{fv}$ 为降水预报 $f$ 与观测 $v$ 的二维离散概率分布， $S_{fv}$ 是对应的评分矩阵。记无技巧预报的评分为 $S_C$ ，“完美预报”评分为 $S_P$ 。若要评分是“公平的”，则式（1）需要满足条件：

$$\text{Perfect FC: } \sum_v P_v S_{vv} = S_P \quad (2)$$

$$\text{Constant FC: } \sum_v P_v S_{fv} = S_C \quad \forall f \quad (3)$$

式中， $P_v$ 为降水观测 $v$ 离散形式的概率分布。

Rodwell等<sup>[14]</sup>基于Gandin等<sup>[16]</sup>的研究，将降水预报检验评分SEEPS定义为概率空间中预报误差的一种度量，满足式（2）和（3），且 $S_C=1$ ， $S_P=0$ 。SEEPS值域期望为 $[0,1]$ ，1代表无技巧预报，0代表“完美预报”。定义SEEPS技巧评分为 $1-\text{SEEPS}$ 。在保证能够具有足够的检验评估能力的前提下，需要将SEEPS评分对抽样不确定性的敏感性最小化。为此，SEEPS还需满足限制条件：

$$\text{Strong Perfect FC: } S_{vv} = 0 \quad \forall v \quad (4)$$

这个限制条件可以使一个“完美预报”的SEEPS评分不仅期望为0，且任一有限样本的平均值也为0。

### 1.3 SEEPS误差矩阵数学模型

Rodwell等<sup>[14]</sup>设计了式（1）中预报误差评分 $S_{fv}$ 的一种结构：

$$S_{fv} = \begin{cases} |f-v|a + \delta_{if}(c-a), & v > f \\ |f-v|b + \delta_{vi}(d-b), & v < f \\ 0, & v = f \end{cases} \quad (5)$$

式中， $a$ 、 $b$ 、 $c$ 和 $d$ 为正值。若 $i=j$ ，则 $\delta_{ij}=1$ ；若 $i \neq j$ ，则 $\delta_{ij}=0$ 。已经证明，若要此误差评分结构满足常量预报限制条件式（3），则需要分类 $n \leq 3$ 。为获得更高的检验分辨能力，选定SEEPS的降水分类数目 $n$ 为3。3分类误差评分的结构形式如表2所示。

表2 SEEPS  $3 \times 3$ 误差评分矩阵结构<sup>[14]</sup>

		观测			
概率		$P_1$	$P_2$	$P_3$	
		$v$			
Cat		1	2	3	
预报	$f$	1	0	$c$	$c+a$
	2	$d$	0	$a$	
	3	$d+b$	$b$	0	

除限制条件式（4）之外，式（2）和式（3）可以写为：

$$\text{Perfect FC: } \sum_v P_v S_{vv} = 0 \quad (6)$$

$$\text{Constant FC: } \sum_v P_v S_{fv} = 1 \quad \forall f \quad (7)$$

利用限制条件式(4)、(6)和(7),可由式(5)获得关系式:

$$\begin{aligned} b &= \frac{P_3 a}{1 - P_3} \\ c &= \frac{1 - P_3 a}{1 - P_1} \\ d &= \frac{1 - P_3 a}{P_1} \end{aligned} \quad (8)$$

此外,考虑一种极限情况以确定 $a$ 的取值。假设存在某种预报,总是能正确预报“干”和“小雨”(或“小雨”和“大雨”),但观测为“大雨”(或“干”)时,却总是预报为“小雨”。由表2误差评分结构可知这两类预报误差评分底限是 $\min(p_3 a, 1 - p_3 a)$ 。当 $a=1/(2p_3)$ 时,评分底限取极大值0.5。 $a$ 取此值,可以奖励预报所有可能出现的结果,而不是只预报发生频率高的分类。

由 $a=1/(2p_3)$ 和关系式(8),以及 $P_1+P_2+P_3=1$ , $P_2=2P_3$ ,可以获得以降水分类“干”的概率 $P_1$ 表达的SEEPS误差评分矩阵:

$$S_{fv} = \frac{1}{2} \begin{bmatrix} 0 & \frac{1}{1-P_1} & \frac{4}{1-P_1} \\ \frac{1}{P_1} & 0 & \frac{3}{1-P_1} \\ \frac{1}{P_1} + \frac{3}{2+P_1} & \frac{3}{2+P_1} & 0 \end{bmatrix} \quad (9)$$

由(9)式可以看到,SEEPS误差评分矩阵由 $P_1$ 唯一确定。在不同气候区域,“干”的概率值 $P_1$ 差别很大,误差评分矩阵差异也很大,见表3。

表3 SEEPS误差评分矩阵(潮湿气候 $P_1=0.1$ ,干旱气候 $P_1=0.85$ )<sup>[14]</sup>

	观测		
	“干”	“小雨”	“大雨”
概率	0.10	0.60	0.30
“干”	0.00	0.56	2.22
“小雨”	5.00	0.00	1.67
“大雨”	5.71	0.71	0.00
概率	0.50	0.33	0.17
“干”	0.00	1.00	4.00
“小雨”	1.00	0.00	3.00
“大雨”	1.60	0.60	0.00
概率	0.85	0.10	0.05
“干”	0.00	3.33	13.33
“小雨”	0.59	0.00	10.00
“大雨”	1.11	0.53	0.00

从表3中可见,对于多雨气候区, $P_1$ 较小,SEEPS误差评分矩阵的下三角元素(对应“干”的漏报)较大;对于干旱气候区域, $P_1$ 较大,SEEPS误差评分矩阵的上三角元素(对应“大雨”的漏报)较大。这说明基于降水发生概率( $1-P_1$ )的SEEPS误差评分矩阵随不同气候区域或降水季节而变;因此SEEPS的突出特点是对不同气候区域或降水季节的同一预报特征(如对“大雨”的漏报)具有不同的误差评分,即对预报误差的“惩罚”与降水气候概率相关,故SEEPS可自动适应不同气候区域或降水季节的站点降水概率,使得多个站点的SEEPS误差评分融合更合理,区域平均和时间平均更有意义。

#### 1.4 SEEPS与降水气候特征相关联的误差评分特点

ECMWF利用观测站点历史降水资料计算SEEPS误差矩阵。因为SEEPS评分中降水概率具有年变化特点,利用质量控制的站点逐月历史降水观测数据,逐月定义“干”的概率阈值 $P_1$ ,换句话说,某一观测站SEEPS误差评分矩阵是逐月变化的,同一个月内每天的SEEPS检验都采用同一个误差评分矩阵;由式(9)可计算得到各观测站点的SEEPS误差评分矩阵。事实上,由于极端降水的频次及降水量值对概率 $P_1$ 影响很小,故SEEPS误差评分矩阵对历史降水数据的质量控制要求并不高。

在SEEPS降水分类中,将“小雨”和“大雨”概率关系设定为 $P_2=2P_3$ ,而非 $P_2=P_3$ 或 $P_2=3P_3$ 的考虑是,既尽可能提高区分“小雨”和“大雨”的阈值,以提高对高量级降水的辨识能力,又不至于使SEEPS误差评分的标准差明显变大,而让SEEPS变得很不稳定。

#### 1.5 SEEPS根据站点密度加权区域平均

针对观测站点检验的最大问题是观测站地理分布不均对评分统计后处理结果代表性的不利影响。如中国区域的降水观测站具有东部密集西部稀疏的特点,如何融合疏密不同地区的SEEPS误差评分,获得有代表性的区域平均检验结果,以避免高站点密度区域的SEEPS值主导区域平均SEEPS结果呢?SEEPS检验方法提出了加权区域平均计算方案。

如检验采用 $N$ 个观测站点,设 $\rho_k$ 为站点 $k$ 附近的站点密度,用高斯核函数表达为:

$$\rho_k = \sum_l e^{-(\alpha_{kl}/\alpha_0)^2} \quad (10)$$

式中, $\alpha_{kl}$ 是站点 $k$ , $l$ 对地心的张角,由这两个站点的地理经纬度唯一确定。 $\alpha_0$ 是一个参考角度,取 $0.75^\circ(83\text{km})$ 。当 $\alpha_{kl}>4\alpha_0$ 时,站点 $l$ 对 $\rho_k$ 计算的影响可以

忽略不计。因为 $a_{kk}=0$ ，所以 $\rho_k \geq 1$ ，设站点权重系数

$$W_k = \frac{1}{\sum_l \rho_l} \rho_k$$

为 $W_k$ ，与站点密度呈反相关，则 $W_k = \frac{\rho_k}{\sum_l \rho_l}$ 。

降水检验误差评分SEEPS的加权区域平均表达式为：

$$S = \sum_k W_k S_k \quad (11)$$

考虑站点分布密度后的SEEPS区域平均值，可以抑制高站点密度区域的SEEPS评分在区域平均SEEPS评分中所占比重，能够更客观地表达预报在整个预报区域内的整体预报能力。

## 2 SEEPS在中央气象台定量降水预报检验中的应用试验

### 2.1 降水观测站及权重分布

观测站点为中国范围内参与GTS国际交换的384个观测站点。观测站点分布东密西疏，特别是在西藏、新疆的大片区域内，站点分布稀少。按照SEEPS方法中的站点权重分布（图略），站点稀疏的西部地区的站点权重，约是站点稠密的东部地区的两倍左右。这可以平抑东部地区稠密站点的SEEPS误差评分在中国区域平均SEEPS误差评分中所占比重，使区域平均SEEPS更能代表预报在全国范围内的整体预报能力。

### 2.2 检验数据

预报数据采用中央气象台2007年3月—2013年12月共7年的7天预报时效内24h定量降水预报（quantitative precipitation forecast, QPF）。由于预报员QPF的站点预报是降水等级区间预报，在无准确的站点预报值的情况下，统计检验中采取的预报匹配方法较为简单。将QPF站点预报值定义为如表4所示的降水量值。

表4 预报员QPF预报的相当降水量（mm）

	QPF预报区间	预报相当值
预报	0	0.0
	0~2	1.0
	2~5	3.5
	5~10	7.5
	10~20	15.0
	20~25	22.5
	25~50	37.5
	50~100	75.0
	100~250	175.0
	≥250	350.0

### 2.3 中国区域 SEEPS 降水分类特征

在对SEEPS检验试验结果进行展示分析之前，先分析SEEPS评分的降水分类特征，可以帮助更深入

地理解SEEPS检验结果。由30年气候降水观测数据（1980—2009年）统计获得的中国区域1, 4, 7和10月的24h累积降水“干”概率分布（图略）可见，冬季干，夏季湿，春季和秋季的湿中心不同，前者位于华南地区，后者位于西南地区的四川盆地地区。7月中国范围内最“干”（ $P_1$ 大）的区域是西北地区，“干”概率 $P_1$ 超过0.8，降水极少。降水频率最高（ $1-P_1$ ）区域在西南地区，包括西藏东南部、四川大部及云南等地区。这里的日降水概率普遍超过0.5，很“湿润”。

从“大雨”的阈值分布（图1）来看，“大雨”阈值整体呈现由东南往西北逐渐递减的特点。该阈值在江淮流域较大，最高达到25~30mm，与传统降水检验方法“大雨”量级相当，说明该地区“大雨”的降水量较高；“大雨”阈值在西北地区较小，只有5~10mm，说明该地区“大雨”的降水量较小。

综合“干”概率分布及“大雨”阈值分布特点，可以知道，西北地区降水概率小、“大雨”阈值很小，说明该地区降水量很小；降水量较大的地区为降水概率大的西南地区及“大雨”阈值高的江淮流域，以及华南地区。

## 3 SEEPS与传统评分的QPF检验对比分析

在具有显著气候差异的干旱少雨地区和热带湿润地区，由于SEEPS降水分类基于降水分类的概率比例，使得SEEPS方法与传统评分具有不同的应用特征。在降水的分类上，传统降水检验评分采取的是固定阈值；如24h累积降水分类中的“无雨”、“小雨”、“中雨”等的量级固定为： $<0.1\text{mm}$ ， $0.1\sim 10\text{mm}$ 和 $10\sim 25\text{mm}$ 等。这种采取固定阈值的方法，从降水气候概率角度看是不合理的，它忽略了不同地区（或季节）的降水气候差异；例如，同一时期的江淮流域降水量明显高于西北地区，而两者用同一暴雨阈值计算检验评分，会使得西北地区对“中雨”及以上量级降水站数 $a$ 的贡献远远低于江淮流域。这种采用固定阈值的做法，会使得高降水气候概率的区域（或季节）主导（如TS、Bias、空报率、漏报率等）与降水分类列联表中 $d$ 无关的所有评分。在SEEPS降水检验方法中，依据观测站点降水气候概率，在概率空间中将该站点降水划分为三类，并使“小雨”发生概率为“大雨”的两倍。这样，区分“小雨”与“大雨”的降水阈值是变化的。变化阈值的优点就是，可以使具有不同降水气候概率的地区，如新疆地区与江淮流域，具有同样机会来影响区域平均SEEPS评分。

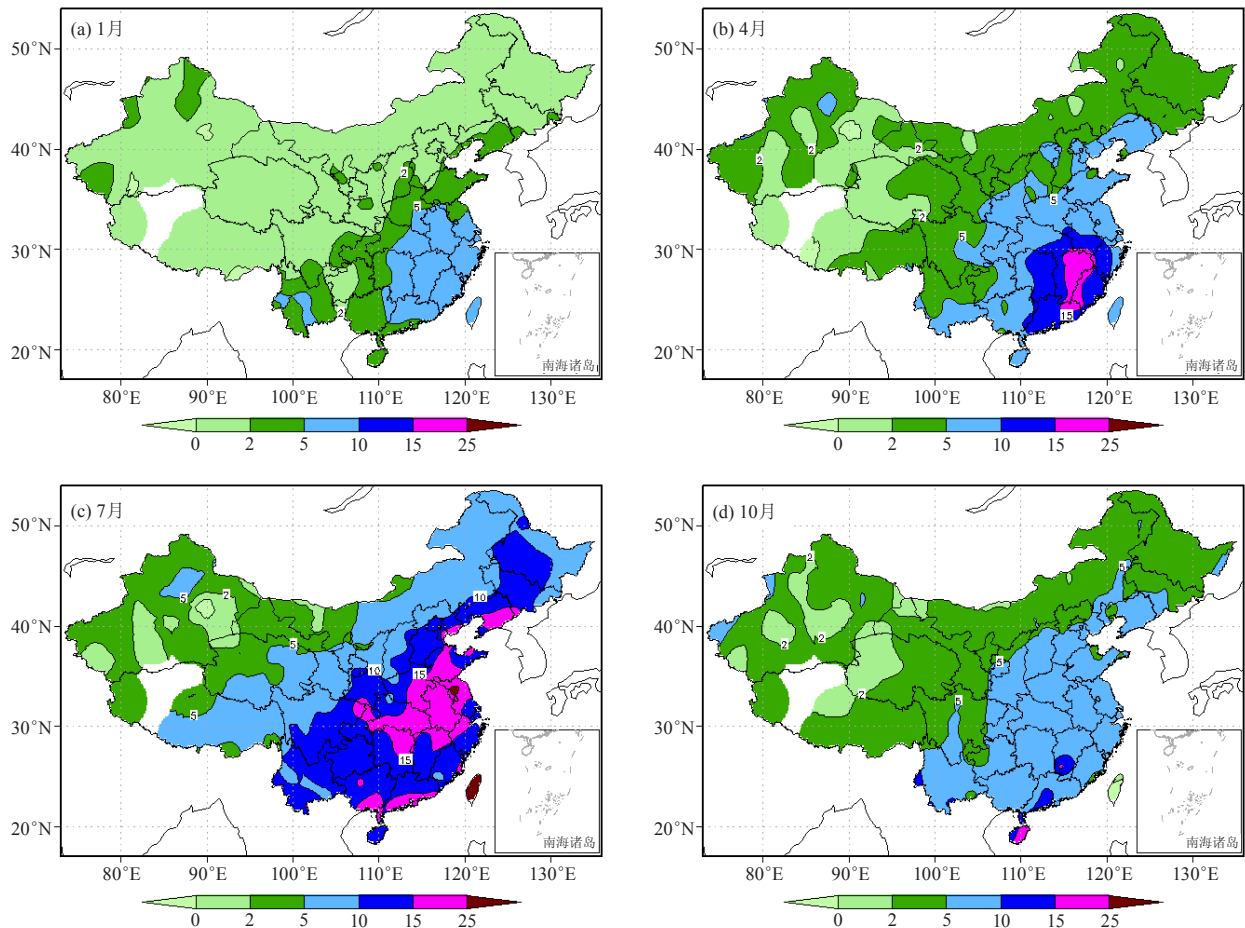


图1 基于SEEPS降水分类的中国区域1 (a), 4 (b), 7 (c) 和10月 (d) 24h累积降水 (00UTC) 的“大雨” 阈值分布

降水的年变化与上述降水的地区差异一样, 也会使传统检验方法的固定阈值与SEEPS方法的变化阈值具有不同的特征。对于传统检验方法, 在年尺度的评分计算中, 固定阈值会使得降水少的“干”季, 对“中雨”及以上量级的降水站数 $a$ 的贡献远小于降水量多的“湿”季, 从而使得“湿”季主导整个评分; 或在逐日评分的时间平均中, 因“中雨”及以上量级在“干”季缺少定义, 而使得“湿”季主导整个评分。在SEEPS方法中, 利用一年中不同时间的降水气候概率, 定义该时间的降水分类, 故每一个观测站点上的降水阈值是随时间变化的。SEEPS方法采取的这种随时间变化的降水阈值, 可以使“干”季与“湿”季在整体平均的SEEPS评分中具有相当的影响力。SEEPS方法是基于降水概率采用的随时间、空间变化的降水阈值, 使得SEEPS评分均值具有更完整的时间和空间代表性。

图2是预报员在中国区域24h累积降水预报的SEEPS技巧评分。从图2可见, 预报员1~7d时效预报SEEPS技巧评分呈现波浪式上升, 2013年达到最高,

2010年评分相对较低。图3是预报员传统检验评分的公平成功指数ETS评分。总体上看, 各量级2013年的ETS评分均达到最大值, 但在其他年份, 不同量级的降水表现不一。以2012年为例, 小雨量级的ETS评分是历年中最低的, 而中雨量级的ETS几乎是历年中较高的, 比较难以给出一个综合性的预报水平的评估。SEEPS评分则不同, 可将不同量级的评分综合成一个参数, 便于综合评估。

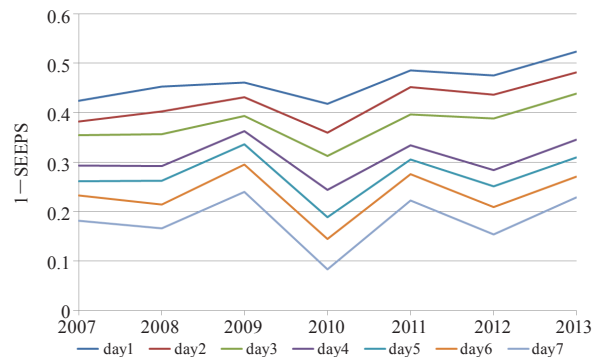


图2 2007—2013年中国区域24h累积降水 (00UTC) QPF 的SEEPS技巧评分

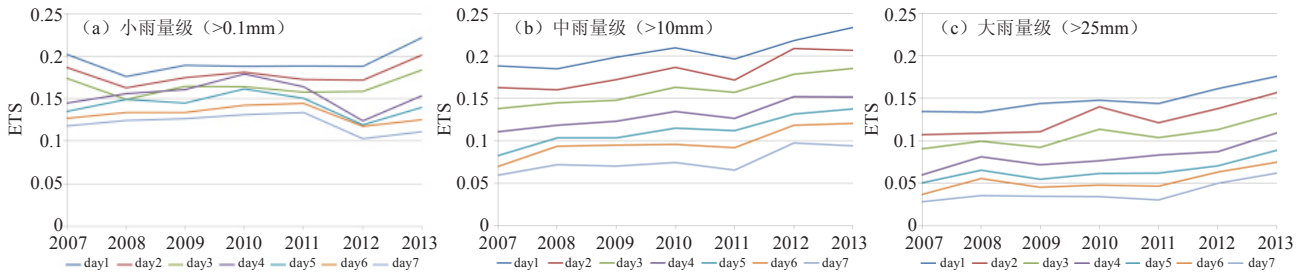


图3 2007—2013年预报员24h降水QPF公平成功指数ETS评分

图4是2013年8月预报员24h累积降水48h时效的QPF预报SEEPS误差。QPF预报SEEPS总误差为0.56，SEEPS技巧评分为0.44。图中展示了各类漏报和空报的空间分布，其中3类漏报的SEEPS误差之和为0.26，3类空报SEEPS误差之和为0.20；这说明漏报、空报对预报效果的影响大致相当。从图中预报员QPF的SEEPS误差分布特征可见，预报“干”对比观测“小雨”和预报“干”对比观测“大雨”这两类漏报的SEEPS误差高值区主要分布在新疆、内蒙古等西北地区；预报“小雨”对比观测“大雨”此种类型漏报

的SEEPS误差高值区主要分布于华北、东北地区。预报“小雨”对比观测“干”这个类型的空报误差分布高值区主要位于西南地区，而预报“大雨”对比观测“干”这个类型的空报误差分布高值区主要位于西南、西北地区。总体上说，预报员QPF预报误差的主要来源为两类：预报“小雨”对比观测“大雨”的漏报和预报“小雨”对比观测“干”类型的空报，合计占到了总误差的65%；前者说明预报员降水预报量级较实际降水偏小，后者说明预报员对“小雨”的大范围、高频率空报也可以导致总体预报效果的明显下降。

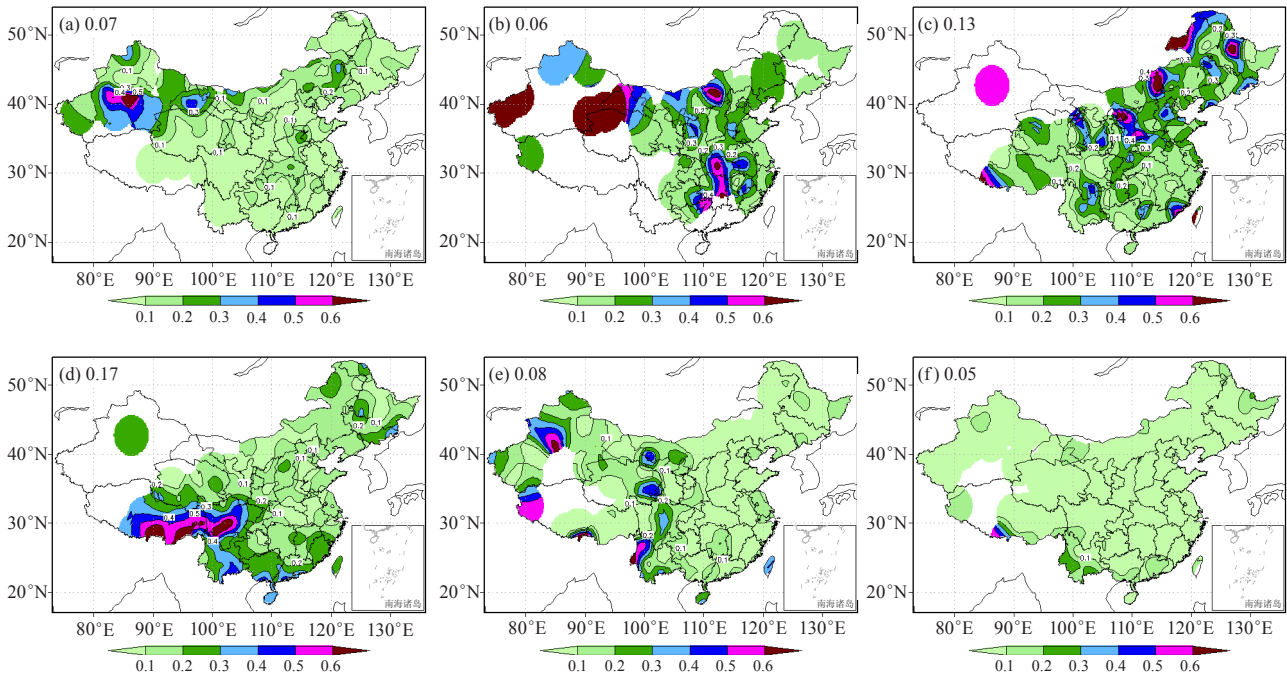


图4 2013年8月预报员24h累积降水48h预报SEEPS误差分类分布,预报对比观测  
(漏报: (a) “干”对比“小雨”; (b) “干”对比“大雨”; (c) “小雨”对比“大雨”;  
空报: (d) “小雨”对比“干”; (e) “大雨”对比“干”; (f) “大雨”对比“小雨”)

#### 4 总结与讨论

降水检验SEEPS方法是近期ECMWF研究的评估降水预报技巧的单一评分，本文将SEEPS方法应用到预报员中国区域QPF降水预报检验中，介绍了基于降

水概率分布的三分类降水误差检验方法SEEPS的计算原理、约束条件和计算数学方案，对中央气象台预报员定量降水预报进行了评估，并与传统检验评分进行了对比分析。

SEEPS方法基于降水气候概率，在概率空间中进行降水分类，使SEEPS评分具有可随时间、空间变化的可变降水阈值，克服了传统检验方法中，因较为显著的空间、时间上的降水概率差异而引起的不同检验评分难以直接对比分析的困难，从而在概率空间中可将预报误差定量化。SEEPS评分具有与站点密度呈反相关的站点权重，使其区域平均值不受高站点密度区域的主导，具有更好的区域整体代表性；逐日SEEPS评分的时间平均值与由检验时段内所有样本计算获得的SEEPS评分等价，这使得SEEPS评分的时间平均具有更明晰的实际意义。SEEPS利用降水气候概率，在概率空间中把降水分为“干”、“小雨”和“大雨”三类，使得“小雨”概率为“大雨”概率的两倍。由此，SEEPS将降水检验评分表达为两个矩阵的标量积：一个 $3 \times 3$ 列联表矩阵，为降水观测与预报的二维离散概率分布，代表了概率空间中的降水预报特征；另一个为基于降水概率的 $3 \times 3$ 误差评分矩阵，代表了误差评分特征。降水预报特征体现了概率空间中的模式（或预报员）预报性能；误差评分特征表现了在一定降水概率下，SEEPS评分对概率空间中不同预报误差的“惩罚”。SEEPS将检验评分分为“预报”与“评分”这两部分，因而可以分析一个预报SEEPS评分高低的成因。SEEPS在不同降水概率下具有不同的误差评分特征，使其能够自动适应不同的降水气候。

SEEPS对降水预报的评分总体上与传统检验评分的结论是相当的，但检验指标更加简单，更加直接，其提供了清晰、简洁的定量化指标，便于管理与决策。从图1可以看到，以SEEPS方法的降水分类获得的“大雨”量级阈值，即使在7月，也只有江淮很小区域达到传统检验评分的“大雨”量级，因此SEEPS方法的不足是对传统降水检验方法中的“大雨”及以上量级的降水预报无检验分辨能力，不适用于对此类大量级降水事件的检验。对于极端气候条件（极“干”， $P_1 > 0.85$ 或极“湿”， $P_1 < 0.1$ ）下的降水预报，SEEPS方法也不适用；对于此类属于小概率事件的降水，其预报检验需要用专门的极端事件检验方法<sup>[17]</sup>。

目前国际上SEEPS方法的应用仅限于24h累积降水的预报检验。进一步拓展SEEPS方法的应用范围，将其用于更高时间分辨率（如6h累积）的降水预报检验是SEEPS方法下一步的重要研究方向之一。

致谢：感谢中国气象局数值预报中心王雨、李莉、李应林、赵滨和邓莲堂等提供的建议与帮助。

#### 参考文献

- [1] 崔锦, 周晓珊, 张爱忠, 等. 天气学检验在东北区域数值模式秋冬季降水预报中的应用. 气象与环境学报, 2009, 25(4): 17-21.
- [2] Wilks D S. Statistical Methods in the Atmospheric Sciences (2nd). New York: Academic Press, 2006.
- [3] Jolliffe I T, Stephenson D B. Forecast Verification: A Practitioner's Guide in Atmospheric Science. New York: Wiley and Sons Ltd, 2003.
- [4] 王雨. 若干数值模式对2003年夏季青藏高原中南部降水预报检验. 高原气象, 2004, 23: 53-58.
- [5] 王雨, 闫之辉. 降水检验方案变化对降水检验评估效果的影响分析. 气象, 2007, 33(12): 53-61.
- [6] 王雨, 公颖, 陈法敬, 等. 区域业务模式6h降水预报检验方案比较. 应用气象学报, 2013, 24(2): 171-178.
- [7] Ebert E E. Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. J Appl Meteor, 2008, 15: 51-64.
- [8] Casati B. New developments of the intensity-scale technique within the Spatial Verification Methods Intercomparison Project. Wea Forecasting, 2010, 25: 113-143.
- [9] Davis C B, Brown B, Bullock R G. Object-based verification of precipitation forecasts, Part I: Methods and application to mesoscale rain areas. Mon Wea Rev, 2006, 134: 1772-1784.
- [10] Davis C A, Brown B, Bullock R G. Object-based verification of precipitation forecasts, Part II: Application to convective rain systems. Mon Wea Rev, 2006, 134: 1785-1795.
- [11] Davis C A, Brown B, Bullock R G, et al. The Method for Object-based Diagnostic Evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC Spring Program. Wea Forecasting, 2009, 24: 1252-1267.
- [12] Richardson D S. Skill and relative economic value of the ECMWF ensemble prediction system. Quart J Roy Meteor Soc, 2000, 126: 649-667.
- [13] Ferro C A T, Stephenson D B. Extremal Dependence Indices: Improved verification measures for deterministic forecasts of rare binary events. Wea Forecasting, 2011, 26: 699-713.
- [14] Rodwell M J, Richardson D S, Hewson T D, et al. A new equitable score suitable for verifying precipitation in numerical weather prediction. Quart J Roy Meteor Soc, 2010, 136: 1344-1363.
- [15] Haiden T M, Rodwell M J, Richardson D S. Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score. Mon Wea Rev, 2012, 140: 2720-2733.
- [16] Gandin L S, Murphy A H. Equitable skill scores for categorical forecasts. Mon Wea Rev, 1992, 120: 361-370.
- [17] Stephenson D B, Casati B, Ferro CAT, et al. The extreme dependency score: A non-vanishing measure for forecasts of rare events. J Appl Meteor, 2008, 15:41-50.