

DOI: 10.13476/j.cnki.nsbdcqk.2019.0003

胡义明,梁忠民,蒋晓蕾,等. 1GFS集合降雨预报的校正后处理研究[J]. 南水北调与水利科技, 2019, 17(1): 15-19. HU Y M, LIANG Z M, JIANG X L, et al. Study on statistical post-processing of GFS ensemble precipitation forecasts[J]. South-to-North Water Transfers and Water Science & Technology, 2019, 17(1): 15-19. (in Chinese)

GFS 集合降雨预报的校正后处理研究

胡义明, 梁忠民, 蒋晓蕾, 王军, 李彬权

(河海大学 水文水资源学院, 南京 210098)

摘要:耦合降水集合预报信息与水文模型是水文预报发展的一个重要方向。然而由于大气运行初始条件及模式的不确定性数值降雨预报不可避免地存在误差。基于全球集合预报系统(GFS)提供的1~8 d 预见期的降雨集合预报数据,研究了基于扩展型 Logistic 算法和异方差扩展型 Logistic 算法发展的5个统计后处理模型对准河流域息县子流域 GFS 预报降雨的校正效果。结果表明:5个模型对 GFS 预报降雨均具有较好的校正效果,但随着预见期的增长,各个模型的校正能力呈衰减趋势。总体而言,相较于基于扩展型 Logistic 算法的3个模型,基于异方差扩展型 Logistic 算法的2个模型具有更优的校正能力。

关键词:集合降雨预报;集合后处理;扩展型 Logistic 算法;异方差扩展型 Logistic 算法;GFS

中图分类号:TV213 文献标志码:A 开放科学(资源服务)标识码(OSID):



Study on statistical post-processing of GFS ensemble precipitation forecasts

HU Yiming, LIANG Zhongmin, JIANG Xiaolei, WANG Jun, LI binquan

(College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China)

Abstract: Coupling ensemble precipitation prediction with hydrological models is an important development direction of hydrological forecasting. However, due to the uncertainty of the initial atmospheric conditions and model physics, numerical precipitation forecasts inevitably have errors. In this study, based on GFS ensemble precipitation reforecasts with a 1-8 day lead time, we analyzed five statistical post-processing models that were developed based on extended logistic regression (ELR) and heteroscedastic extended logistic regression (HELRL) algorithms, and compared their correction effects on the GFS ensemble precipitation forecasts in Xixian sub-basin of Huai River basin. The results indicated that: these five models all made significant improvements to the GFS raw forecast; but with the extension of the lead time, their correction effects tended to attenuate. In general, the two HELRL-based models had better performance compared to the other three ELR-based models.

Key words: ensemble precipitation forecast; ensemble post-processing; extended logistic algorithm; heteroscedastic extended logistic algorithm; GFS

利用数值天气预报产品驱动水文模型进行径流预报是提高预报精度、增长预见期的有效途径^[1-3]。目前世界上很多气象机构提供从小时到周及至月年

尺度不同预见期的数值预报产品,如中尺度天气预报欧洲中心 ECMWF、美国国家海洋和大气管理局的环境预测中心、中国国家气候中心等发布的数值

收稿日期:2018-07-15 修回日期:2018-11-20 网络出版时间:2018-12-05

网络出版地址: <http://kns.cnki.net/kcms/detail/13.1334.TV.20181130.1119.002.html>

基金项目:国家自然科学基金(41730750;51709073);江苏省自然科学基金(BK20170878)

作者简介:胡义明(1986—),男,江苏宿迁人,讲师,博士,主要从事水文水资源方面研究。E-mail: yiming.hu@hhu.edu.cn1

预报产品。但由于大气系统的高度非线性,加之预报模型及模型输入等不确定性因素的存在,无论哪种产品都存在预报误差及集合预报低离散度问题,所以需要降雨集合预报产品进行有效地校正处理,以提高其精度和可靠性^[4-6]。

近十年来,围绕如何通过统计后处理技术来提高天气/气候预报的精度和可靠性这一问题,开展了大量卓有成效的研究并已取得了一些代表性的研究成果。如, Raftery 等^[7]采用贝叶斯模型平均法(BMA)后处理集合气温预报,其中,每个集合成员的预报系列都采用正态分布函数进行拟合。考虑到降雨的偏态分布及降雨系列中存在大量零值情况, Sloughter 等^[8]采用逻辑回归函数和伽马分布函数分别描述 0 值降雨事件和非 0 值降雨事件的分布,进而通过构建混合分布函数描述降雨事件的分布,对 BMA 方法进行了改进,使其能应用到集合降雨预报的后处理。 Hamill 等^[9]采用逻辑回归方法建立降雨预报分布与集合预报均值间的统计关系去后处理集合降雨预报,获得在给定集合降雨预报条件下“真实降雨”的条件分布函数。考虑到在传统逻辑回归方法中,回归方程的拟合依赖于阈值的选取,即不同的阈值,对应着不同的回归方程曲线,且不同阈值对应的回归方程曲线可能存在交叉情况, Wilks^[10]对传统逻辑回归进行了改进,提出了将阈值(分位点)作为自变量引入到方程,进而可提供连续的概率分布函数去描述“真实降雨”的条件概率分布。

本文采用基于扩展型 Logistic 算法和异方差扩展型 Logistic 算法而发展的 5 个统计后处理模型,以淮河流域息县子流域的降雨预报为对象,研究 5 个不同的后处理模型对全球集合预报系统(Global ensemble forecast system, GFS)提供的降雨集合预报产品的校正能力。

1 Logistic 方法基本原理

Logistic 回归方法属于非线性的统计回归方法,标准的 Logistic 方法的研究对象是离散的二分类事件,即 $Y=1$ 或 $Y=0$ 。二分类事件 Y 在自变量 X 影响下的发生概率的计算可采用式(1),也即标准的 Logistic 模型

$$p(y \leq q | \mathbf{X}) = \frac{\exp[\mathbf{X}^T \mathbf{B}]}{1 + \exp[\mathbf{X}^T \mathbf{B}]} \quad (1)$$

式中: \mathbf{X} 表示自变量组成的矩阵; q 是给定的阈值(分位点),用以决定二分类事件 Y 的发生与否($Y=1$ 或 $Y=0$); $\mathbf{B} = \{B_0, B_1, B_2, \dots\}$ 是参数矩阵,包括截距及各变量的回归系数,参数 \mathbf{B} 的估计是通过最

大化对数似然函数获得的,即

$$\ln L = \operatorname{argmax}_{\mathbf{B}} \sum_{i=1}^N \ln(\pi_i) \quad (2)$$

式中: N 是观测样本系列长度; π_i 是第 i 个观测值对应的概率,其计算为

$$\pi_i = \begin{cases} P(y_i \leq q | x_i, \mathbf{B}), & y_i \leq q \\ 1 - P(y_i \leq q | x_i, \mathbf{B}), & y_i \geq q \end{cases} \quad (3)$$

对于不同的阈值 q , Logistic 回归系数是不一样的,这就有可能导致不同阈值条件下的回归方程出现交叉情况,而使得对于某些值而言,求出的概率为无效的负概率值,即如果 $p(y \leq q_a | \mathbf{X}) > p(y \leq q_b | \mathbf{X})$ 时,就会出现 $p(q_a \leq y \leq q_b) < 0$ 情况。

为了解决上述问题及尽量使 Logistic 回归方程中具有更少的参数,扩展的 Logistic 回归方法被提出^[10]。在扩展的 Logistic 模型中,分位点 q (阈值)被当作一个自变量引入到方程中,从而使得 Logistic 方法可以提供全概率分布用于描述因变量 Y ,而事件 Y 不必是二分类事件,即^[10]

$$p(y < q | \mathbf{X}) = \frac{\exp[\mathbf{X}^T \mathbf{B} + g(q)]}{1 + \exp[\mathbf{X}^T \mathbf{B} + g(q)]} \quad (4)$$

式中: $g(q)$ 是分位点 q 的函数。

在扩展的 Logistic 方法基础上,异方差扩展的 Logistic 方法又被提出^[11]。相比较于扩展的 Logistic 方法仅能调整 Y 条件概率分布 $p(y < q | \mathbf{X})$ 的位置,异方差扩展的 Logistic 方法能够同时对因变量 Y 条件概率分布的位置和离散度进行调整。该模型可描述为^[11]

$$p(g(y) < g(q) | \mathbf{X}) = \frac{\exp\left[\frac{g(q) - \mu}{\sigma}\right]}{1 + \exp\left[\frac{g(q) - \mu}{\sigma}\right]} \quad (5)$$

式中: μ 是分布的位置参数; σ 是分布的尺度参数,具体表示为

$$\mu = \mathbf{X}^T \boldsymbol{\omega}, \sigma = \exp(\mathbf{H}^T \boldsymbol{\lambda}) \quad (6)$$

式中: \mathbf{X} 和 \mathbf{H} 表示自变量矩阵; $\boldsymbol{\omega}$ 和 $\boldsymbol{\lambda}$ 回归系数。

关于公式(4)和(5)中的函数 $g(q)$, 研究表明^[10]: $g(q) = a\sqrt{q}$ 可以使得上述 2 个方法达到较好的精度。为此,本研究中,分位点(阈值) q 的函数采用式(7)

$$g(q) = a\sqrt{q} \quad (7)$$

基于扩展的 Logistic 方法(式(4))和异方差扩展的 Logistic 方法(式(5)),设计了 5 套不同的校正模型用于处理集合预报降雨数据。

第一个模型(M1)表示为

$$p(y \leq q) = \frac{\exp(aM + b\sqrt{q} + c)}{1 + \exp(aM + b\sqrt{q} + c)} \quad (8)$$

式中: M 表示集合预报 \sqrt{X} 的均值; a, b, c 是模型参数; q 是分位点; p 是不超过概率($y \leq q$)。

第二个模型(M2)表示为

$$p(y \leq q) = \frac{\exp(aM + bS + c\sqrt{q} + d)}{1 + \exp(aM + bS + c\sqrt{q} + d)} \quad (9)$$

式中: S 表示集合预报 \sqrt{X} 的标准差; a, b, c, d 是模型参数;其它参数同上。

第三个模型(M3)表示为

$$p(x \leq q) = \frac{\exp[(aM + b(M * S) + c\sqrt{q} + d)]}{1 + \exp[(aM + b(M * S) + c\sqrt{q} + d)]} \quad (10)$$

式中: a, b, c, d 是模型参数;其它参数同上。

第四个模型(M4)表示为

$$p(x \leq q) = \frac{\exp\left[\left(\frac{a\sqrt{q} - bM + c}{\exp(dS)}\right)\right]}{1 + \exp\left[\left(\frac{a\sqrt{q} - bM + c}{\exp(dS)}\right)\right]} \quad (11)$$

式中: a, b, c, d 是模型参数;其它参数同上。

第五个模型(M5)表示为

$$p(x \leq q) = \frac{\exp\left(\frac{a\sqrt{q} - (bM + cS) + d}{\exp(hS)}\right)}{1 + \exp\left(\frac{a\sqrt{q} - (bM + cS) + d}{\exp(hS)}\right)} \quad (12)$$

式中: a, b, c, d, h 是模型参数;其它参数同上。

在上述 5 个校正模型(M1 至 M5)中,M1、M2 和 M3 模型是基于扩展的 Logistic 方法(式(4))设计的,而 M4 和 M5 模型是基于异方差扩展的 Logistic 方法(式(5))演变而来的。基于历史观测和同期的集合预报降雨数据,对 5 个模型的校正效果进行评估,最终选取最优模型对未来的集合预报降雨数据进行校正处理。

2 实例研究

以淮河流域上游的息县子流域未来 8 d 预报降雨数据为研究对象,采用上文介绍的 5 种后处理模型对息县流域未来 8 d 降雨预报进行校正处理。实测降雨数据为 2006—2009 年共 4 年,未来 8 d 的降雨预报数据来源于 GFS 预报系统。GFS 全球集合预报系统是由美国国家环境预报中心研制,系统具有 11 个集合预报成员,其回预报数据从 1984 年 12 月至今。GFS 每天发布 4 次(0:00、6:00、12:00 和 18:00 UTC)气象预报数据,滚动免费提供未来 8~16 d 的降雨预报数据,其中 8 d 预见期内数据的时段间隔为 6 h,降雨数据的空间分辨率为 Gaussian~0.5°。基于 4 层交叉验证思路,分析了不同模型的校正效果,即每次采用 3 年降雨数据率定模型参数,

剩余的 1 年数据用以验证。

图 1 给出了基于赤池信息量准则 AIC 指标^[12]评估的 5 个模型的校正效果。AIC 因其可综合考虑模型的简洁性(参数个数)和精确性,是一种常用地衡量统计模型拟合优良性的指标,其值越小,模型精度越高。从图中可以看出,就第 1 d 至第 5 d 预见期而言,第 5 模型(M5)是最优模型,第 4 模型(M4)次之;对于第 6 d 至第 8 d 预见期而言,5 个模型性能差异性不大。

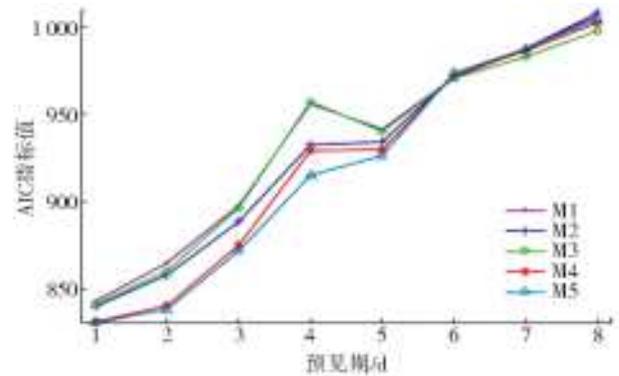


图 1 不同后处理校正模型校正效果的 AIC 评估指标值

Fig.1 AIC statistic values for the five different post-processing models

图 2 给出了原始预报和经 5 个模型校正后预报的秩概率技巧得分(ranked probability skill score, RPSS),RPSS 值越大表明预报效果越优^[6]。从图中可以看出,GFS 原始预报在前 3 d 预见期内呈现一定的预报技巧,但从第 4 d 开始,技巧得分为负数,即相对于气候预报而言,GFS 原始预报亦没有提高预报精度。总体而言,经 5 个模型校正后预报的技巧得分均高于 GFS 原始预报,且技巧得分均为正值,说明模型具有较好的校正效果。就 5 个模型的校正效果而言,在第 1 d 至第 5 d 预见期内,第 5 模型(M5)的校正效果最优,第 4 模型(M4)效果次之。而在第 6 d 至第 8 d 预见期,5 个模型的精度差异性不大。这与基于 AIC 指标的评估结果一致。

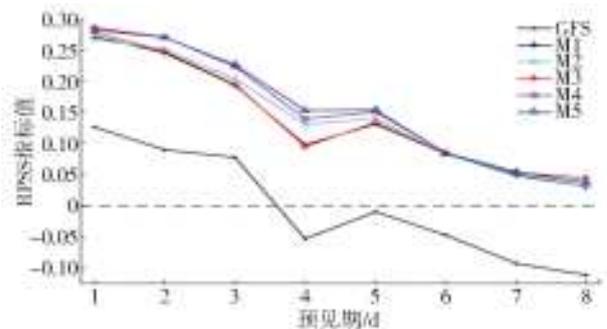


图 2 不同后处理校正模型校正效果的 RPSS 技巧得分

Fig.2 Ranked probability skill score (RPSS) for GFS raw forecast and the five post-processing models

进一步详细分析第 5 模型 M5 对 GFS 原始预报的校正效果,在预见期为 1 d 和 2 d 情形下预报结果的校正效果,在预见期为 1 d 和 2 d 情形下预报结果见图 3。对超过 5 mm 降雨事件预报的可靠性曲线(reliability diagram),可靠性曲线越接近 1:1 直

线,预报可靠性越高^[13]。相比于 GFS 原始预报的可靠性曲线而言,经 M5 模型校正后预报的可靠性曲线更接近与 1:1 直线,表明经模型 M5 后处理后,预报的可靠性可明显改善。

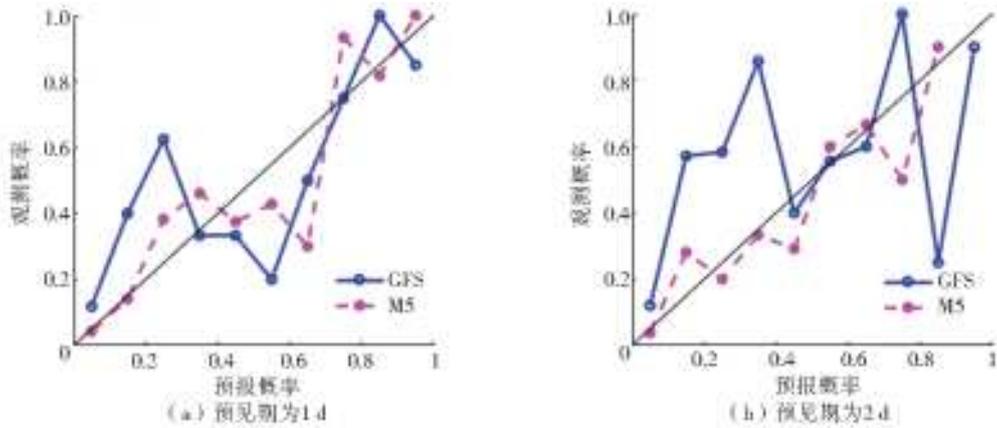


图 3 GFS 原始预报和 M5 模型校正效果的可靠性曲线
Fig. 3 Reliability diagrams of the GFS raw forecast and M5

图 4 给出在预见期为 1 d 和 2 d 情形下,基于 GFS 原始预报和经 M5 模型校正后预报。可以看出,对超过 5 mm 降雨事件预报的相对作用特征曲线(relative operating characteristic, ROC),ROC 曲线越向左上方靠近或其包围的曲线下方面积越大,

表明预报效果越优^[13]。相比于 GFS 原始预报的 ROC 曲线,经 M5 模型校正后预报的 ROC 曲线更靠近左上角,且曲线下方包围的面积也比 GFS 预报的 ROC 曲线包围的面积大,这都表明经模型 M5 后处理后,预报的命中率有较大提高。

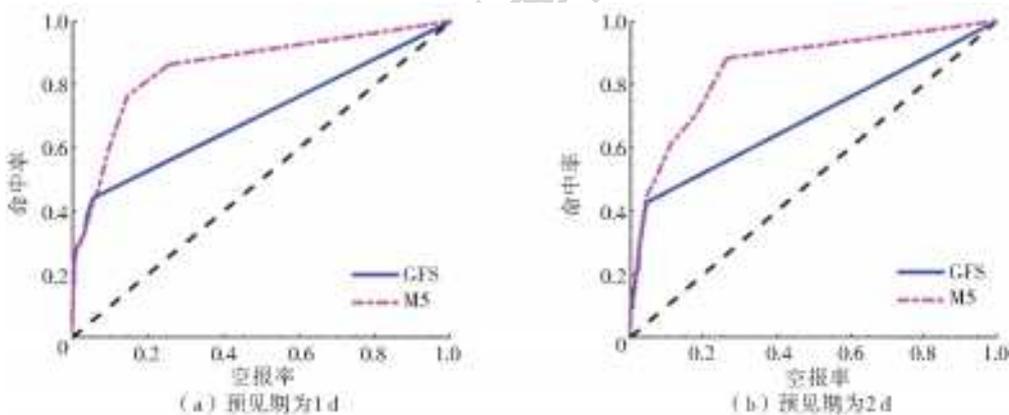


图 4 GFS 原始预报和 M5 模型校正效果的 ROC 曲线
Fig. 4 Relative operating characteristics (ROC) diagrams of GFS raw forecast and M5

3 结论

本文采用扩展型 Logistic 算法和异方差扩展型 Logistic 算法衍生的 5 个统计后处理模型,以淮河流域息县子流域为例,研究了 5 个不同统计后处理模型对 GFS 降雨集合预报产品的校正能力。结果表明,5 个模型对 GFS 预报均具有较好的校正效果,但随着预见期的增长,各个模型的校正能力呈衰减趋势。对于第 1~5 d 的预见期而言,第 4 模型 M4 和第 5 模型 M5 效果较优,其中 M5 稍优于 M4;对于第 6~7 d 而言,5 个模型的效果相当;而对于第 8 d 而言,M3 和 M4 模型较好。不同的预

见期,各模型的校正效果有所差别,但总体而言, M5 模型效果最优。这主要由于在 M5 模型中,模型的位置参数同时考虑了集合预报的均值和方差的影响,而模型的尺度参数考虑了集合预报方差的影响,这使得在不同的预报时刻,可根据集合预报的均值和方差,对预报分布函数的位置和形状进行实时调整。

参考文献(References):

[1] 彭勇,王萍,徐炜,等. 气象集合预报的研究进展[J]. 南水北调与水利科技, 2012, 10(4): 90-96. (PENG Y, WANG P, XU W, et al. Review of weather ensemble

- prediction [J]. *South-to-North Water Diversion and Water Science & Technology*, 2012, 10(4): 90-96. (in Chinese))
- [2] 徐静,叶爱中,毛玉娜,等. 水文集合预报研究与应用综述[J]. *南水北调与水利科技*, 2014, 12(1): 82-87. XU J, YE A Z, MAO Y N, et al. Review of research and application of hydrologic ensemble forecast [J]. *South-to-North Water Transfers and Water Science and Technology*, 2014, 12(1): 82-87. (in Chinese))
- [3] 赵琳娜,刘莹,党皓飞,等. 集合数值预报在洪水预报中的应用进展[J]. *应用气象学报*, 2014, 25(6): 641-653. (ZHAO L N, LIU Y, DANG H F, et al. The progress on application of ensemble prediction to flood forecasting [J]. *Journal of Applied Meteorological Science*, 2014, 25(6): 641-653. (in Chinese))
- [4] 叶爱中,段青云,徐静,等. 基于 GFS 的飞来峡流域水文集合预报[J]. *气象科技进展*, 2015, 5(3): 57-61. (YE A Z, DUAN Q Y, XU J, et al. The hydrological ensemble prediction based on GFS in Feilaixia basin [J]. *South-to-North Water Transfers and Water Science and Technology*, 2015, 5(3): 57-61. (in Chinese))
- [5] 杜雅玲,陆桂华,吴志勇,等. 基于 TIGGE 数据的中期降水预报在江苏省的应用效果评估[J]. *南水北调与水利科技*, 2017, 15(6): 17-25. (DU Y L, LU G H, WU Z Y, et al. Assessment of TIGGE medium-term precipitation forecast in Jiangsu province [J]. *South-to-North Water Transfers and Water Science and Technology*, 2017, 15(6): 17-25 (in Chinese))
- [6] HU Y M, SCHMEITS M J, JAN VAN ANDEL S, et al. A stratified sampling approach for improved sampling from a calibrated ensemble forecast distribution [J]. *Journal of Hydrometeorology*, 2016, 17(9): 2405-2417.
- [7] RAFTERY A E, GNEITING T, BALABDAOUI F, et al. Using Bayesian model averaging to calibrate forecast ensembles [J]. *Monthly Weather Review*, 2005, 133(5): 1155-1174.
- [8] SLOUGHTER J M, RAFTERY A E, GNEITING T, et al. Probabilistic quantitative precipitation forecasting using Bayesian model averaging [J]. *Monthly Weather Review*, 2007, 135(9): 3209-3220.
- [9] HAMILL T M, WHITAKER J S, WEI X. Ensemble reforecasting: improving medium-range forecast skill using retrospective forecasts [J]. *Monthly Weather Review*, 2004, 132(6): 1434-1447.
- [10] WILKS D S. Extending logistic regression to provide full probability distribution MOS forecasts [J]. *Meteorological Applications*, 2009, 16(3): 361-368.
- [11] MESSNER J W, MARY G J, ZEILEIS A, et al. Heteroscedastic extended logistic regression for Postprocessing of ensemble guidance [J]. *Monthly Weather Review*, 2014, 142(1): 448-456
- [12] HU Y M, LIANG Z M, CHEN X, et al. Estimation of design flood using EWT and ENE metrics and uncertainty analysis under non-stationary conditions [J]. *Stochastic Environmental Research and Risk Assessment*, 2017, 31(10): 2617-2626.
- [13] WILKS D S. *Statistical methods in the atmospheric sciences-third edition* [M]. The Netherlands Elsevier, Amsterdam, 2011.

《南水北调与水利科技》期刊亮点之加入科技部“OSID 计划”

本刊已加入科技部“OSID 计划”(Open Science Identity 开放科学计划),该计划通过在期刊每篇文章上添加 OSID 开放科学(资源服务)标识码,为读者提供增值服务。作者需要将本文的 1 分钟语音介绍及开放的科学内容与数据等补充材料上传到 OSID 码中。有利于科研成果交流传播,提升期刊创新能力,扩大论文与期刊的影响力,建立学术评价标准。