

DOI:10.20040/j.cnki.1000-7709.2023.20221327

# 基于 GAMLSS 模型的云南省 非一致性标准化降水指数分析

陈 灏<sup>1</sup>, 张晓琳<sup>1</sup>, 胡彩霞<sup>1</sup>, 张艳敏<sup>2</sup>, 陈 畅<sup>1</sup>, 董前进<sup>3</sup>

(1. 珠江水文水资源勘测中心, 广东 广州 510370; 2. 中工武大设计集团有限公司, 湖北 武汉 430070;  
3. 武汉大学水资源与水电工程科学国家重点实验室, 湖北 武汉 430072)

**摘要:** 干旱指数对干旱的研究至关重要。随着气候变化和人类活动的影响,水文序列的频率分布随时间变化,因此需考虑干旱指数的非一致性。标准化降水指数  $S_{SPI}$  认为降水服从 Gamma 分布,对应的分布函数参数通常稳定不变,而实际的分布函数参数与时间序列具有一定相关性,因此利用 GAMLSS 模型以时间序列为协变量构建了降水序列的非一致性 Gamma 分布模型,在此基础上计算出非一致性标准化降水指数  $N_{NSPI}$ 。以云南省为例,分别计算了 1960~2019 年的逐月  $N_{NSPI}$ 、 $S_{SPI}$ ,将两者进行对比分析,发现两者对干旱的识别结果总体趋势相近,但在 1960~1985 年  $N_{NSPI}$  识别的干旱数量和等级强于  $S_{SPI}$ ,1986~2005 年两者相近,而在 2006~2019 年,  $S_{SPI}$  识别的干旱数量和等级逐渐超过  $N_{NSPI}$ ,进一步分析表明,考虑了分布函数参数随时间变化的  $N_{NSPI}$  能更精确反映实际干旱情况。

**关键词:** 干旱指数; NSPI; GAMLSS 模型; 云南省

**中图分类号:** TV213.4

**文献标志码:** A

**文章编号:** 1000-7709(2023)04-0009-04

## 1 概况

云南省位于我国西南边陲,总面积约  $39.4 \times 10^4 \text{ km}^2$ ,其中山地面积占全省总面积的 85%,平均海拔约为 2 000 m<sup>[1]</sup>。云南省基本属于亚热带高原季风型气候,年温差较小,日温差较大,干湿季节分明,年平均降水量约 1 100 mm<sup>[2]</sup>,湿季(5~10 月)降水量约占全年的 85%,干季(11 月~次年 4 月)降水量约占全年 15%<sup>[3]</sup>。考虑干旱指数的非一致性,对于准确反映区域实际干旱情况有重要意义。本文以云南省为例,利用 GAMLSS 模型构建降水序列随时间变化的非一致性 Gamma 分布模型,并在此基础上计算出该区域的非一致性干旱指数  $N_{NSPI}$ ,以分析讨论该区域的干旱情况。

## 2 数据与方法

### 2.1 研究数据

选取云南省及其周边共 30 个气象站点(图 1)

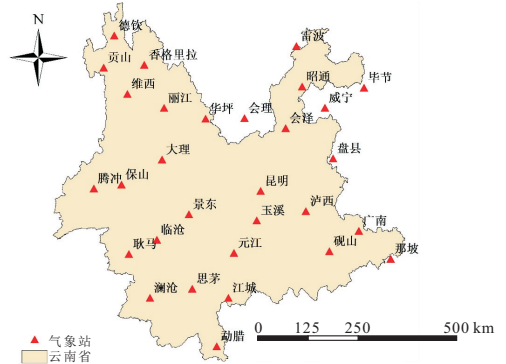


图 1 云南省主要气象站点分布图

Fig. 1 Distribution map of main meteorological stations in Yunnan Province

1960~2019 年共 60 年的逐日降水数据,数据来源于中国气象数据共享网(<https://data.cma.cn/>),由日降水量数据计算得到各站点逐月降水量数据,将各站逐月数据进行泰森多边形处理后得到云南省 1960~2019 年的逐月降水量数据。

为检验数据一致性,利用 Mann-Kendall 检验方法<sup>[4]</sup>对 30 个气象站及云南省的逐月降水数据进行突变性检验,发现各站点均存在突变,突变

收稿日期: 2022-06-28,修回日期: 2022-08-01

基金项目: 国家自然科学基金项目(51979198)

作者简介: 陈灏(1994-),男,硕士、工程师,研究方向为水资源规划与管理,E-mail:chenhaoqy@whu.edu.cn

通讯作者: 董前进(1979-),男,博士、副教授,研究方向为水资源规划与管理,E-mail:dqjin@whu.edu.cn

开始时间各不相同。综合各站计算所得的云南省逐月降水数据突变从 2004 年 8 月开始,初步检验可知降水数据存在非一致性。

## 2.2 研究方法

### 2.2.1 GAMLSS 模型

GAMLSS 模型假设事件  $y_i (i = 1, 2, \dots, n)$  相互独立,  $f(y_i | \theta^i)$  为  $y_i$  对应的概率密度分布函数,其中  $\theta^i = (\theta_1^i, \theta_2^i, \theta_3^i, \theta_4^i)$  为概率分布函数所对应的参数,每个参数都可以是解释变量的函数,定义一个单调函数  $g_k(\cdot) (k = 1, 2, 3, 4)$  分别作为各分布函数参数的连接函数,则 GAMLSS 模型可定义为:

$$g_k(\theta_k) = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(x_{jk}) \quad (1)$$

式中,  $\theta_k$  为长度为  $n$  的向量,表示第  $k$  个参数对应  $y_i (i = 1, 2, \dots, n)$  概率分布函数的取值;  $\mathbf{X}_k$  为解释变量,为  $n \times m$  矩阵;  $\boldsymbol{\beta}_k$  为长度为  $m$  的参数向量;  $h_{jk}(\cdot)$  为关于解释变量  $x_{jk}$  的函数 ( $j = 1, 2, \dots, J_k$ ), 可以是线性或非线性函数。

利用 GAMLSS 构建云南省非一致性降水分布模型,通过 R 语言 gamlss 包实现,假设降水随时间序列服从两参数的 Gamma 分布,则对应的 GAMLSS 模型包含 Location 参数  $\mu$  及 Scale 参数  $\sigma$ ,令 Location 参数  $\mu$  作为时间序列  $t$  的函数,即:

$$g_1(\mu) = g_1[\mu(t)] = b_0 + b_1 t + \dots + b_m t^m \quad (2)$$

式中,  $b_m (m = 0, 1, \dots, n)$  为多项式系数;  $m$  为多项式次数,根据 AIC 最小原则确定拟合曲线。

确定拟合曲线后将需计算的时刻代入式(2)即可得到相应时刻的分布函数参数,从而确定相应时刻的降水分布函数。

### 2.2.2 NSPI

(1) 标准化降水指数  $S_{SPI}$ 。利用随机变量  $x$  表示某降水序列,则其 Gamma 分布的概率密度函数为:

$$f(x) = \frac{1}{\beta^\gamma \Gamma(\gamma)} x^{\gamma-1} e^{-x/\beta} \quad x > 0 \quad (3)$$

$$\Gamma(\gamma) = \int_0^\infty x^{\gamma-1} e^{-x} dx \quad (4)$$

式中,  $\beta$  为尺度参数;  $\gamma$  为形状参数。

确定式(3)中各相关参数后,对于降水样本中某一确定的降水量  $x_0$ ,可通过对式(3)进行数值积分以求出其累计概率  $G(x_0)$ :

$$G(x_0) = \int_0^{x_0} f(x) dx = \int_0^{x_0} \frac{1}{\beta^\gamma \Gamma(\gamma)} x^{\gamma-1} e^{-x/\beta} dx \quad (5)$$

由于降水量可能出现为零的情况,但式(5)不

能准确计算  $x_0 = 0$  的概率,因此改进式(5)以使其可计算  $x_0 = 0$  的情况,具体表达式为:

$$H(x_0) = q + (1 - q)G(x_0) \quad (6)$$

其中  $q = m/n$

式中,  $H(x_0)$  为降水累积概率;  $q$  为降水量为 0 的概率;  $m$  为降水量为 0 的样本数;  $n$  为总样本数。

将式(6)求得的  $H(x_0)$  代入标准化正态分布函数进行正态标准化处理即可得到  $S_{SPI}$ ,即:

$$H(x_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx \quad (7)$$

(2) 非一致性干旱指数  $N_{NSPI}$ 。利用非一致性 GAMLSS 模型计算出降水累积频率,同样代入式(7)进行正态标准化处理,即可得到  $N_{NSPI}$  指数,由于两者均服从 Gamma 分布,且均利用降水累积频率通过正态标准化而得,因此  $N_{NSPI}$  及  $S_{SPI}$  对干旱的分级相同,见表 1<sup>[5]</sup>。

表 1  $N_{NSPI}$  及  $S_{SPI}$  对干旱等级的划分

| 序号 | 干旱类型 | 干旱等级  | 范围                                  |
|----|------|-------|-------------------------------------|
| 1  | 无旱   | $D_0$ | $-0.5 < N_{NSPI}/S_{SPI}$           |
| 2  | 轻旱   | $D_1$ | $-1.0 < N_{NSPI}/S_{SPI} \leq -0.5$ |
| 3  | 中旱   | $D_2$ | $-1.5 < N_{NSPI}/S_{SPI} \leq -1.0$ |
| 4  | 重旱   | $D_3$ | $-2.0 < N_{NSPI}/S_{SPI} \leq -1.5$ |
| 5  | 特旱   | $D_4$ | $N_{NSPI}/S_{SPI} \leq -2.0$        |

## 3 结果与讨论

### 3.1 结果与分析

将云南省 1960~2019 年逐月降水量按照 1~12 月分别构建以时间序列为协变量的 GAMLSS 模型,按照 AIC 最小的原则选择  $g_1(\mu)$  的曲线拟合方式(利用 R 语言 gamlss 包中平滑函数 pb() 自动拟合最优曲线),最终确定的最优各月的模型参数见表 2。由表 2 可知,1~12 月的最优模型均能够达到“均值接近 0,方差接近 1,偏态系数接近 0,峰态系数接近 3,Filliben 系数大于等于 0.977”的标准<sup>[6]</sup>,因此认为各月份的 GAMLSS 模型构建合理。

图 2 为 1、4、7、10 月份最优 GAMLSS 模型的残差 worm 图,从图 2 中可以看出偏差值较为均匀地分布在 95% 置信区间内(黑色虚线范围内),无明显超出部分,因此 Gamma 分布的残差拟合满足条件。

为进一步讨论 GAMLSS 模型的优势,考虑分布函数参数随时间序列线性变化的 GLM 模型,GLM 模型与 GAMLSS 模型的降水分布函数与参数相同。计算发现,GAMLSS 模型所拟合的

表 2 各月的模型参数结果及残差特征值

Tab. 2 Results of model parameters and residual eigenvalues in each month

| 月份 | 位置参数     |          | 尺度参数     | 均值       | 方差      | 偏态系数     | 峰态系数    | Filleben 系数 |
|----|----------|----------|----------|----------|---------|----------|---------|-------------|
|    | 常数项      | 拟合系数     |          |          |         |          |         |             |
| 1  | -7.615 2 | 0.004 8  | -0.334 4 | 0.001 9  | 1.018 0 | -0.134 8 | 2.717 6 | 0.993 9     |
| 2  | 9.146 4  | -0.003 6 | -0.464 1 | 0.008 0  | 1.021 2 | -0.671 6 | 2.966 7 | 0.977 3     |
| 3  | -3.700 7 | 0.003 1  | -0.668 1 | -0.001 3 | 1.016 3 | 0.137 9  | 2.373 7 | 0.993 2     |
| 4  | 0.861 9  | 0.001 0  | -0.934 7 | -0.000 4 | 1.016 8 | 0.054 7  | 2.261 4 | 0.992 9     |
| 5  | 7.344 7  | -0.001 9 | -0.970 0 | 0.001 6  | 1.017 4 | -0.401 2 | 2.382 8 | 0.986 3     |
| 6  | 12.920 2 | -0.004 4 | -1.642 4 | 0.000 0  | 1.016 9 | 0.015 9  | 2.112 8 | 0.991 8     |
| 7  | 5.488 8  | -0.000 5 | -1.956 2 | 0.000 1  | 1.017 0 | -0.158 4 | 3.112 0 | 0.992 0     |
| 8  | 10.897 0 | -0.003 3 | -1.655 6 | 0.000 3  | 1.017 0 | -0.324 5 | 3.203 4 | 0.992 9     |
| 9  | 7.865 8  | -0.002 0 | -1.692 2 | 0.000 1  | 1.017 0 | -0.148 9 | 2.973 7 | 0.987 0     |
| 10 | 13.870 4 | -0.005 2 | -1.021 9 | -0.000 7 | 1.016 7 | 0.146 0  | 2.093 1 | 0.990 7     |
| 11 | 37.780 1 | -0.017 6 | -0.371 0 | 0.004 5  | 1.019 5 | -0.324 7 | 2.687 4 | 0.991 2     |
| 12 | 27.242 0 | -0.012 8 | -0.165 0 | -0.007 4 | 1.010 7 | 0.317 9  | 2.530 8 | 0.990 6     |

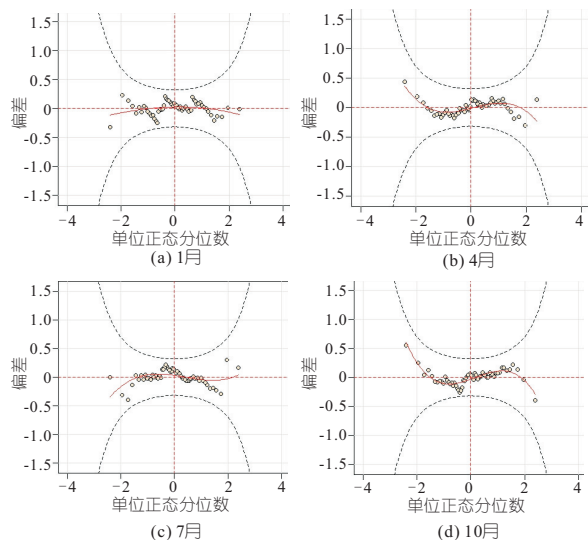


图 2 1、4、7、10 月份残差 worm 图

Fig. 2 Worm plots in January, April, July, and October

Gamma 分布在降水累积频率  $P$  较小的情况下, 相同的降水量所对应的理论频率要小于 GLM 模型(如图 3 以 2019 年为例, 各月参数取平均值), 即 GAMLSS 模型所识别的干旱要比 GLM 严重, 因此 GAMLSS 模型对干旱的识别更为灵敏。

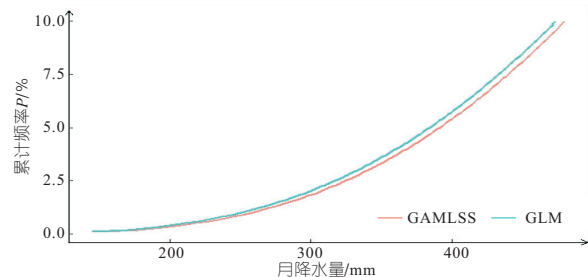


图 3 2019 年 GAMLSS 和 GLM 模型累积频率曲线对比图  
Fig. 3 Comparison of cumulative frequency curves of GAMLSS and GLM models in 2019

另外, GAMLSS 模型考虑了最优的参数拟合曲线, 其 AIC 指数比 GLM 模型要小, 参数拟合的更好, 因此选择 GAMLSS 模型构建降水分布的非一致性模型。

利用 1~12 月的 GAMLSS 模型分别计算每

个月降水的累积概率, 通过式(7)计算出  $N_{NSPI}$  指数, 将各月  $N_{NSPI}$  指数按照时间序列整合, 同时根据式(3)~(7)计算出相应时间的  $S_{SPI}$ , 为便于对比分析, 将无旱的月份指数 ( $-0.5 < N_{NSPI}/S_{SPI}$ ) 标记为 0, 将干旱月份指数 ( $N_{NSPI}/S_{SPI} \leq -0.5$ ) 全部取为相反数, 将处理后的  $N_{NSPI}$ 、 $S_{SPI}$  绘制成折线图(图 4)进行对比分析。

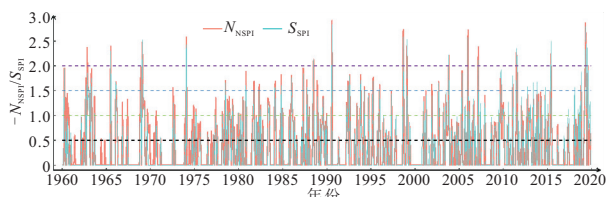


图 4  $N_{NSPI}$ 、 $S_{SPI}$  对比分析图

Fig. 4 Comparison analysis plot of  $N_{NSPI}$  and  $S_{SPI}$

在图 4 中利用辅助线区分 4 种干旱级别的范围。由图 4 可知,  $N_{NSPI}$ 、 $S_{SPI}$  的整体趋势大致相同, 说明  $N_{NSPI}$  的计算相对合理, 但两者存在一定差异, 在 1960~1985 年间,  $N_{NSPI}$  对干旱的识别无论是数量还是等级均明显超过  $S_{SPI}$ , 1986~2005 年间,  $N_{NSPI}$ 、 $S_{SPI}$  对干旱的识别较为接近, 2006~2019 年,  $S_{SPI}$  对干旱的识别数量及其等级逐渐超过  $N_{NSPI}$ 。由于  $N_{NSPI}$  考虑了分布函数随时间的变化, 1960~1985 年, 降水减少更加容易达到干旱出现的频率, 随着时间序列的延长, 同样的少量降水在整体降水量排序中逐渐向中间靠拢, 因此在 1960~1985 年  $N_{NSPI}$  更容易识别出干旱, 而在 1985~2005 年  $N_{NSPI}$ 、 $S_{SPI}$  对干旱的识别情况较为接近, 在 2006~2019 年  $S_{SPI}$  对干旱的识别逐渐超过  $N_{NSPI}$ 。由于时间序列总是逐渐延长,  $N_{NSPI}$  在考虑降水的概率分布时, 能够结合每个时间点的最新降水资料修正概率分布参数, 相对于  $S_{SPI}$  能更合理地分析干旱情况。

为更直观分析  $N_{NSPI}$ 、 $S_{SPI}$  对干旱的识别情况, 将两者对不同等级干旱的识别结果绘制成柱状图, 见图 5。由图 5 可知, 1960~2019 年 60 年

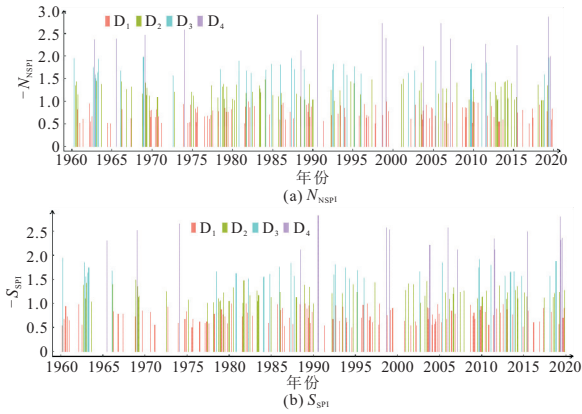


图 5  $N_{NSPI}$ 、 $S_{SPI}$  不同等级干旱分布图

Fig. 5 Drought distribution plots of  $N_{NSPI}$  and  $S_{SPI}$  grades

间,  $N_{NSPI}$  在 1960~1985 年对干旱的识别数量和等级均明显超过  $S_{SPI}$ , 在 1986~2005 年两者对干旱识别的结果较为接近, 而在 2006~2019 年,  $N_{NSPI}$  对干旱的识别等级逐渐低于  $S_{SPI}$ 。可见  $N_{NSPI}$  考虑了降水分布函数参数随时间变化, 相比于  $S_{SPI}$  能更合理、准确地识别干旱的发生。

### 3.2 讨论

选择合适的干旱指数对干旱评估与监测至关重要, 传统的  $S_{SPI}$  是以降水序列服从一致性 Gamma 分布为基础的, 然而由于气候变化和人类活动的影响, 较长时间的降水序列往往不再服从一致性概率分布<sup>[7]</sup>, 从而降低了  $S_{SPI}$  对干旱评估的准确性, 因此有必要构建降水的非一致性概率分布, 并据此计算出非一致性  $S_{SPI}$ 。

本研究中利用 CAMLSS 模型构建了降水的非一致性 Gamma 分布, 更符合实际降水概率分布情况, 特别是随着降水序列的延长, 受环境变化的影响越多, 非一致性干旱指数  $N_{NSPI}$  相较于传统的  $S_{SPI}$  更能适用于干旱评估。

然而, 本次非一致性概率分布模型仅考虑以时间为协变量, 由于气候变化及人类活动对降水

序列的影响较为复杂, 降水序列的非一致性影响因素较多, 故仅以时间为协变量的  $N_{NSPI}$  对干旱进行评估时可能无法完全反应降水序列的非一致性。在未来的研究中应进一步考虑将气候指数或人为指数作为协变量深入分析。

## 4 结论

a. 由于降水序列随着时间增长总是不断延长,  $N_{NSPI}$  在考虑降水概率分布时结合了不同时刻降水序列的分布特征, 考虑了分布函数参数随时间的变化情况, 相对于  $S_{SPI}$  仅考虑整体的降水序列服从固定参数的 Gamma 分布更具有合理性。

b.  $N_{NSPI}$  相比于传统的  $S_{SPI}$  在识别干旱时具有一定优势, 且  $N_{NSPI}$  与  $S_{SPI}$  的计算原理相似, 具有相同的干旱分级, 因此可在不同区域代替  $S_{SPI}$  进一步推广应用。

### 参考文献:

- [1] 许玲燕, 王慧敏, 段琪彩, 等. 基于 SPEI 的云南省夏玉米生长季干旱时空特征分析[J]. 资源科学, 2013, 35(5): 1024-1034.
- [2] 杨晓静, 左德鹏, 徐宗学. 基于标准化降水指数的云南省近 55 年旱涝演变特征[J]. 资源科学, 2014, 36(3): 473-480.
- [3] 刘小刚, 冷险险, 孙光照, 等. 基于 1961—2100 年 SPI 和 SPEI 的云南省干旱特征评估[J]. 农业机械学报, 2018, 49(12): 236-245, 299.
- [4] 郭明, 杨志勇, 周政辉, 等. 非一致性条件下嘉陵江上游水文频率分析[J]. 华北水利水电大学学报(自然科学版), 2018, 39(1): 42-47.
- [5] LU E. Determining the start, duration, and strength of flood and drought with daily precipitation: Rationale [J]. Geophysical research letters, 2009, 36(12). DOI:10.1029/2009GL038817.
- [6] 陈伏龙, 杨宽, 蔡文静, 等. 基于 GAMLSS 模型的水文干旱指数研究——以玛纳斯河流域为例[J]. 地理研究, 2021, 40(9): 2670-2683.
- [7] 雷晓辉, 王浩, 廖卫红, 等. 变化环境下气象水文预报研究进展[J]. 水利学报, 2018, 49(1): 9-18.

## Analysis of Nonstationary Standardized Precipitation Index in Yunnan Province Based on GAMLSS Model

CHEN Hao<sup>1</sup>, ZHANG Xiao-lin<sup>1</sup>, HU Cai-xia<sup>1</sup>, ZHANG Yan-min<sup>2</sup>, CHEN Chang<sup>1</sup>, DONG Qian-jin<sup>3</sup>

(1. Pearl River Hydrology and Water Resources Survey Center, Guangzhou 510370, China; 2. CEWUD Group Company Limited, Wuhan 430070, China; 3. State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China)

**Abstract:** Drought index is crucial to the study of drought. With the influence of climate change and human activities, the frequency distribution of hydrological time series changes with time, so the nonstationary of drought index needs to be considered. The standardized precipitation index ( $S_{SPI}$ ) considers that the precipitation obeys the Gamma distribution, the corresponding distribution function parameters are usually stationary and invariable. The actual distribution function parameters have a certain correlation with the time series. Therefore, based on the GAMLSS model, the nonstationary Gamma distribution model of precipitation series was constructed with time series as covariate. The nonstationary standardized precipitation index  $N_{NSPI}$  was calculated. Taking Yunnan Province as an example, the monthly  $N_{NSPI}$  and  $S_{SPI}$  during 1960 to 2019 were calculated respectively. Comparing the two indices, it is found that the overall trend of drought recognition results of the two indices is similar. However, in 1960~1985, the number and grade of droughts identified by  $N_{NSPI}$  were stronger than those identified by  $S_{SPI}$ , and the two were similar in 1986~2005. In 2006~2019, the number and grade of droughts identified by  $S_{SPI}$  gradually exceeded those identified by  $S_{SPI}$ . Further analysis shows that the  $N_{NSPI}$  with considering the time-varying distribution function parameters can more accurately reflect the actual drought situation.

**Key words:** drought Index;  $N_{NSPI}$ ; GAMLSS model; Yunnan Province