

文章编号: 1001-6791 (1999) 04-0458-06

非参数统计方法在水文水资源中的应用与展望

王文圣, 丁 晶, 邓育仁

(四川大学水利系, 四川 成都 610065)

摘要: 非参数统计方法在水文水资源系统中的应用和研究已取得了一些进展, 依研究内容不同可分三类: (1) 在水文频率分析中的应用和研究; (2) 在水文水资源预报中的应用和研究; (3) 在水文水资源模拟中的应用和研究。对非参数统计方法在水文水资源系统中的应用和研究现状作了较全面的介绍, 并展望了今后的研究趋势和发展方向。

关 键 词: 非参数统计方法; 水文水资源系统; 核密度估计; 最近邻估计

中图分类号: P333.6 **文献标识码:** A

随着非参数统计理论在数学上的突破和进展, 非参数统计方法在水文水资源系统中的应用和研究也逐渐开展起来。水文水资源系统以降水为输入, 径流为输出, 中间包括蒸发、产流、汇流等复杂过程, 还受到地形、地貌、流域下垫面和人类活动等因素影响^[19]; 同时, 该系统具有随机性、模糊性、灰色性等多种不确定性^[20], 因此它是一个十分复杂的高度非线性的不确定性系统。对这样的系统, 用传统的参数统计模型(建立在假定条件和经验公式上的)描述是不足的, 不能“真实”反映水文水资源系统的客观规律, 而且模型选择具有主观性。非参数统计方法为解决这类问题提供了强有力的工具, 克服参数统计模型的不足。应用研究表明, 非参数统计方法在水文频率分析、水文水资源预报与模拟等方面有比较成功的应用且具有巨大的潜力。

本文对非参数统计方法在水文水资源系统中的应用和研究现状作了较全面的介绍, 并展望了今后的研究趋势和发展方向。

1 非参数统计方法在水文水资源系统中的应用现状

非参数统计方法在水文水资源系统中的应用和研究主要表现在水文频率分析、水文水资源预报与模拟等方面。

1.1 非参数统计方法在水文频率分析中的应用

非参数统计方法为水文频率分析提供了又一条研究途径, 它避开了水文频率计算中困惑多年的线型问题。设 $X_1, X_2, \dots, X_n$ 为来自未知密度函数 $f(X)$ 的独立同分布(i.i.d)样本, 如年最大洪峰流量。对给定频率 p 的设计值 $\hat{X}_p$ 可由下式决定:

收稿日期: 1998-01-20; 修改稿日期: 1998-04-26

基金项目: 自然科学基金资助项目 (49571017)。

作者简介: 王文圣 (1970-) 男, 四川宣汉人, 四川大学水利系, 主要从事水文水资源、水环境新理论、新方法方面研究。

$$p = P\{x \geq \hat{X}_p\} = \int_{\hat{X}_p}^{\infty} \hat{f}(X) dx \quad (1)$$

式中 $\hat{f}(X)$ 为 $f(X)$ 的非参数密度估计, 它对总体分布未作任何假定。显然, 这是一个推求设计值的非参数统计方法。

目前有许多非参数密度估计方法, 如直方图法、Rosenblatt 法、最近邻估计法、核估计法等, 但理论研究最深入、最常用的是核估计法。下面列出核密度估计表达式。单变量核密度函数定义为

$$\hat{f}(X) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{X - X_j}{h}\right) \quad (2)$$

式中 $\hat{f}(X)$ 为总体密度 $f(X)$ 的一个核估计; $K(\cdot)$ 为核函数, 是一概率密度, $\int_{-\infty}^{+\infty} K(X) dX = 1$; h 为带宽; n 为样本容量。多变量核密度函数定义为

$$\hat{f}(X) = \frac{1}{nh^d \det(s)^{1/2}} \sum_{j=1}^n K\left[\frac{(X - X_j)^T S^{-1} (X - X_j)}{h^2}\right] \quad (3)$$

式中 S 是 $X = \{X_1, X_2, \dots, X_d\}^T$ 的样本 $X_i = (X_{i1}, X_{i2}, \dots, X_{id})^T$ ($i = 1, 2, \dots, n$) 协方差矩阵; d 为向量维数; 其余意义同上。

Tung 和 May (1981) 首次将非参数统计方法——自展法 (Bootstrap) 应用于洪水频率分析, 并得出结论: 该法为估算参数的样本方差开辟了一种新途径, 且适合于分析洪水现象的不确定性。K. Adamowski (1985)^[1]首次将非参数核密度估计方法应用于洪水频率分析。该文以实测洪水资料为算例, 非参数核密度估计法推求的洪水设计值与参数统计模型 (三参数对数正态分布 (LN3), 对数皮尔逊III型分布 (LP III), 极值分布 (EV)) 推求的结果一致; 另外, 作者将洪水总体假定为 LP III 分布, 根据总体模拟容量为 50 000 的样本, 由样本用上述参数模型和非参数核密度估计法估计重现期为 50 年和 100 年的洪水设计值。统计试验表明, 非参数核密度估计法较上述参数模型具有小的偏差和均方差。最后作者指出: 该法能描述洪水的多峰模式, 可用于设计洪水。E. Schuster 和 S. Yakowitz (1985)^[2]在洪水频率分析中考虑到常规方法要面临模型选择的困境和非参数密度估计法不能充分利用前期信息、地区信息的缺点, 提出了一种参数与非参数混合的密度估计方法并证明了这种方法的有效性。该法不需要对洪水频率曲线的线型作假定, 能利用前期信息、经验和地区信息, 通过统计试验表明这种途径是令人满意的。1987 年 L. Carles 和 A. Kay 也将非参数核密度估计法应用于水文频率分析中, 取得了较好的结果。1989 年 K. Adamowski^[3]根据实测年洪水资料将非参数核密度估计法与参数模型就不同重现期洪水设计值进行了对比研究, 作者在该文中使用固定核和变化核两种不同的核密度函数, 通过 Monte Carlo 试验得出以下结论: 非参数核密度估计法适合于各种分布的洪水资料, 能给出可靠的洪水设计值; 变核估计法比固定核估计法更具有适用性。1990 年夏乐天引用非参数核密度估计结合再抽样的方法进行洪水频率分析研究^[4], 通过大量模拟试验表明非参数方法的稳健性最好, 其计算结果与各总体的较优参数估计方法的计算结果近似。1991 年郭生练^[5]在考虑我国实测资料不太长而历史洪水和古洪水研究又很有特色的情况下, 提出了考虑历史洪水和古洪水的非参数核密度估计法。通过实例与参数统计模型 (LP III/MOM (矩法), P III/MOM, EV/ML (极大似然法)) 进行了对比研究, 得到了以下结论: (1) 在实测资料区域, 非参数核密度估计法比参数统计模型拟合得更好。(2) 模拟试验表明, 非参数核密度估计法的描述能力和预测能

力都优于参数统计模型。

1.2 非参数统计方法在水文水资源预报中的应用

这里涉及到非参数回归问题。Stone 在 1977 年提出一种非参数回归估计的权函数法。设有来自于 (X, Y) 的随机样本 $(X_i, Y_i) (i = 1, 2, \dots, n)$, 其回归函数 $E(Y/X)$ 的权函数法估计为:

$$\hat{E}(Y/X) = \sum_{i=1}^n W_{ni}(X) Y_i \quad (4)$$

$\sum_{i=1}^n W_{ni}(X) = 1, W_{ni}(X) \geq 0, 1 \leq i \leq n$ 。变量 X 与 Y (均可为多维) 有某种相关关系, 如降雨与径流。构造权 $W_{ni}(X)$ 有两种方法^[6]: 最近邻权法和核权法。最近邻权法中 $W_{ni}(X)$ 构造原理是, 根据样本 $X_i (i = 1, 2, \dots, n)$ 与 X 接近程度 (以欧氏距离判断) 赋不同的权 W_{ni} (其和为 1), X 与 X_i 越接近, W_{ni} 越大。实际应用中 W_{ni} 有很大的选择余地。核权法中 $W_{ni}(X)$ 定义为 (一维)

$$W_{ni}(X) = K \left(\frac{X - X_i}{h} \right) / \sum_{i=1}^n K \left(\frac{X - X_i}{h} \right), \quad i = 1, 2, \dots, n \quad (5)$$

水文水资源非参数统计预报就是基于以上原理进行的。

1985 年 M. Karlsson 在他的博士论文中提出了一种新的降雨径流预报模型——最近邻非参数回归方法 (NN)。S. Yakowitz 1987 年^[7]系统地研究了 NN 法的性质, 这为在水文水资源预报上的应用提供了理论基础, 同时与水文学概念模型进行了对比研究, 结果表明 NN 法是优越的。S. Yakowitz (1985)^[8]将 NN 法应用于洪水警报问题, 并与 ARMA 类模型作了比较。文中以 Kootenai River 1913~1933 年 5 月至 8 月日流量为例进行拟合, 对 1934 年作预报, 计算结果表明: 在均方误差标准下, NN 法与 ARMA (1, 1, 2) 模型拟合得一样好, 对于预报, 前者优于后者; 该文还以 ARMA (3, 2) 模型为总体生成一系列, 然后用 NN 法和 ARMA 模型拟合, 两者的效果基本一致; 作者最后将 NN 法应用于洪水警报问题并取得了有意义的结果。M. Karlsson 和 S. Yakowitz 1987 年将 NN 法用于降雨径流预报问题^[9], 同时还选择了常用参数模型 (ARMAX 模型和瞬时单位线法) 进行了对比, 结果表明 ARMAX 与 NN 法效果接近且满意, 瞬时单位线法较差; 论文还指出, 在样本资料增加到一定时, NN 法保证可收敛到最佳预报值, 而其它参数模型并不具备这个性质。S. Yakowitz 和 M. Karlsson (1987)^[10]应用 Bird Creek 和 Coshocto 两流域的降雨径流资料, 再次把 NN 法同萨克门托模型、单位线法和 ARMAX 模型进行了比较。作者指出 NN 法优于上述模型, 特别适合于洪峰预报。文献^[11]设计了一种新的最近邻核函数为基础的非参数回归方法, 作者认为在最近邻点中某点被抽取的概率服从 Poisson

分布, 取 k 个最近邻点, 则在 k 个最近邻点中第 j 点被抽取的权重为 $W_j = 1/j / \sum_{i=1}^k 1/i, j = 1, 2, \dots, k$ 。赵永龙 1997 年用文献^[11]提出的最近邻核函数构造了相空间近邻点核函数法^[12]并用于岷江紫坪铺站洪水相空间预测。可以看出, 该法优于普通多元回归法。

文献^[13]提出了核权非参数回归方法并对长持续径流——日流量进行了预报研究。该法用于 Washington 行政区用水管理, 取得了较好的预测效果。目前核权非参数回归方法用于水文水资源预报还不多见。

1.3 非参数统计方法在水文水资源模拟中的应用

水文水资源系统的运行与管理都要涉及到该系统各要素的模拟, 即水文水资源时间序列的模拟。对独立时间序列非参数模拟来说, 研究得比较多且取得了一些进展和成果的, 主要有

Bootstrap (自展法) 和 Jackknife (刀切法) 两种方法^[17], 最近文献[18]从贝叶斯观点出发构造了非参数贝叶斯模拟方法, 并用于独立同分布序列模拟研究, 取得了有价值的结果。本文主要介绍相依时间序列非参数统计方法模拟的进展情况。设有相依时间序列 $\{X_i\}$ ($i=1, 2, \dots, n$), (X_i 可以是多维), X_t 取决前 P 个值 $X_{t-1}, X_{t-2}, \dots, X_{t-P}$, 则 X_t 的条件密度函数为

$$f(X_t/X_{t-1}, X_{t-2}, \dots, X_{t-P}) = \frac{f(X_t, X_{t-1}, \dots, X_{t-P})}{\int f(X_t, X_{t-1}, \dots, X_{t-P}) dX_t} \quad (6)$$

非参数统计方法用于模拟的关键就是对式 (6) 作非参数密度估计。

文献[11]提出最近邻非参数自展模式并首次将它用于水文相依时间序列模拟。为了说明这种方法的有效性, 作者使用三个事例作了验证: 以AR(1)和SETAR(1)分别为总体生成100个容量为500的样本。根据样本由最近邻非参数自展模式抽样, 统计试验表明该法能保持总体的线性或非线性相依结构, 统计特性也保持得很好。将该法用于Webber River的月径流模拟并与AR(1)模型模拟结果对比发现最近邻自展模式效果好并能优良地保持分布的偏态和多峰形态。但该法最大缺点是对已知序列的重复抽样, 没有实现合理的内插和外延。

文献[14]构造了非参数核估计干、湿长度(以日为单位, 有雨、无雨的长度)模型并用于单站日降雨量的模拟, 通过与传统的干、湿长度模型对比表明, 非参数核估计模型是优良的, 可用于降雨量模拟。文献[15]建议了非参数核估计一阶马尔柯夫模型(NP(1)), 作者使用该模型分别对AR(1)和SETAR(1)模拟的序列进行统计试验, 结果表明NP(1)能逼近资料的真实分布, 能保持总体的线性或非线性相依关系, 且统计特性保持得良好; 同时将NP(1)用于Beaver River月径流模拟并与AR(1)模型模拟结果对比, 表明NP(1)是优于AR(1), 并且实现了合理的内插和外延, 但该文的工作局限于单变量一阶情况。B. Rajaopalan et al (1997)^[16]构造了以高斯函数为核的多变量一阶非参数核估计模型并给出了详细的算法。通过Salt Lake City 30年日气候资料(降雨量、太阳辐射、最高气温、最低气温、平均露点温度和平均风速)用上述方法对后五种气候因子进行模拟研究。文中降雨量是通过非参数核估计干、湿长度模式^[14]模拟的。Monte Carlo 试验表明, 这种非参数统计方法用于多变量相依时间序列模拟是成功的, 并指出, 带宽 h 的选择和径流超边界现象(如出现径流小于0的情况)是非参数核估计模型所面临的两个急需解决的问题。

2 非参数统计方法在水文水资源系统中的应用研究展望

非参数统计方法不必假定总体分布, 直接从已知样本去推求所需的東西, 与参数统计模型比具有良好的优越性。从目前在水文水资源系统中的应用研究现状来看, 非参数统计方法无论是从独立与相依, 还是从水文频率分析、预报、模拟等方面都取得了一定的进展。但总的来说, 非参数统计方法在水文水资源系统中的应用研究还处于初步阶段, 有许多问题亟待研究。当前宜在以下几个方面展开工作。

(1) 要进一步对非参数统计理论作深入研究。目前非参数统计理论有很大的进展, 但还不完善。比如在进行洪水模拟和预报时, 如何把地区信息和历史洪水信息纳入密度函数估计中; 又如带宽 h 的选择, 虽然有一些方法和原则, 但还不够深入和精确; 核函数 $K(\cdot)$ 也需要引入多种形式进行对比; 还有, 如何把水文水资源中的模糊性、灰色性等不确定性信息耦合进非参数统计模式中(贝叶斯理论可以派上用场); 另外, 非参数统计方法对研究对象如洪水频率曲线和

水文水资源模拟的外延还存在着许多不足;等等。

(2) 要认清参数统计模型与非参数统计模式在水文水资源系统中的地位和作用,要辩证地看待这两种方法,它们各有优势。因此,在认真细致地将非参数统计方法应用于水文水资源系统规划、管理和设计中并与参数统计模型对比,考查各自的优缺点和使用条件及范围。是两种模式并存,取长补短,还是用一种模式取而代之,这是一个非常重要且长期的应用研究工作。我们认为要具体情况具体分析,当参数统计模型拟合得好时就不必考虑非参数统计方法,反之则需考虑;或者结合两者的优点采用混合法描述。

(3) 在水文水资源系统已有的应用研究中,对相依时间序列的模拟和预报,目前仅局限于单变量和多变量一阶非参数统计模式。今后除继续进行单变量和多变量一阶非参数统计模式的应用研究外,还要加强单、多变量多阶非参数统计模式的应用研究。笔者正致力于这方面的研究工作并取得了一定的成果。将非参数统计方法尝试于水文水资源系统中包括模拟、预报、洪水频率分析在内的各个领域,使理论与实践结合起来,既推动了非参数理论的发展,又开辟了水文水资源系统研究的新天地。

参考文献:

- [1] Adamowski K, Nonparametric kernel estimation of flood frequencies[J]. Water Resources Research 1985, 21(11): 1585~ 1590
- [2] Schuster E and Yakowitz S Parametric/Nonparametric mixture density estimation with application to flood- frequency analysis[J]. Water Resources Bulletin 1985, 21(5): 797-814
- [3] Adamowski K A Monte Carlo comparison of parametric and nonparametric estimation of flood frequencies[J], Journal of Hydrology, 1989, 108: 295~ 308
- [4] 夏乐天 水文频率计算中的一种非参数估计方法——密度函数估计法[M]. 2000 年中国水文展望 南京: 河海大学出版社, 1991
- [5] Shi L Guo. Nonparametric variable kernel estimation with historical floods and paleoflood information[J]. Water Resources Research 1991, 27(1): 91~ 98
- [6] 陈希儒, 柴根象编著. 非参数统计教程[M]. 上海: 华东师范大学出版社, 1993
- [7] Yakowitz S. Nearest neighbor methods for time series analysis[J]. J. Time Ser. Anal, 1987, 8(2): 235~ 247.
- [8] Yakowitz S, Markov flow models and the flood warning problem [J]. Water Resources Research 1985, 21(1): 81~ 88
- [9] Karlsson, M and Yakowitz, S. Nearest-neighbor methods for nonparametric rainfall-runoff forecasting[J]. Water Resources Research 1987, 23(7): 1300~ 1308
- [10] Yakowitz S, Karlsson M. Nearest neighbor methods for time series, with application to rainfall/runoff prediction [M]. Stochastic Hydrology edited by J. B. Macneil and G. J. Humphries 1987, 149~ 160
- [11] Upmanu Lal et al. A nearest neighbor bootstrap for resampling hydrologic time series[J], Water Resources Research, 1996, 32(3): 679~ 693
- [12] 赵永龙 水文动力学系统混沌分析及其非线性预测[D]. 四川联合大学博士论文, 1997
- [13] Smith, J A. Long-range streamflow forecasting using nonparametric regression[J]. Water Resources Bulletin 1991, 27(1): 39~ 46
- [14] Upmanu Lal et al. A nonparametric wet/dry spell model for resampling daily precipitation

- [J]. Water Resources Research. 1996, 32(9): 2803~ 2823
- [15] A. Shama et al Stream simulation: a nonparametric approach [J]. Water Resources Research. 1997, 33(2): 291~ 308
- [16] B. Rajagopalan et al Multivariate nonparametric resampling scheme for generation of daily weather variable[M]. Stochastic Hydrology and Hydraulics. 1997: 11: 65~ 93
- [17] Efron, B. Bootstrap methods: Another look at the Jackknife[J], Ann., Stat., 1979, 7: 1~ 26
- [18] Fortin V., et al Simulation, Bayes, and bootstrap in statistical hydrology [J]. Water Resources Research. 1997. 33(3): 439~ 448
- [19] 叶守泽, 夏军编 水文系统识别(原理与方法) [M]. 北京, 水利电力出版社, 1989
- [20] 夏军等编 现代水科学不确定性研究与进展 全国首届水文水资源与水环境中的不确定性研究新理论新方法学术讨论会论文集[C]. 成都: 成都科技大学出版社, 1994

Non-Parametric Statistic Approach: Application and Prospect in Hydrology and Water Resource

WANG Wen-sheng, DENG Jing, DENG Yu-ren

(Sichuan University. Chengdu 610065, China)

Abstract: Some advancements of application and studies of nonparametric statistic approach have been made in the system of hydrology and water resources. Those include three aspects: (1) hydrology frequency analysis, (2) forecasting of hydrology and water resources, and (3) simulating of hydrology and water resources. In this paper, the state of applications and studies of nonparametric statistic approach have been introduced. Finally, some trend and directions for future studies are presented.

Key words: nonparametric statistic approach; system of hydrology and water resources; kernel density estimation; nearest neighbor estimation