

DOI:10.3876/j.issn.1000-1980.2019.02.002

基于多目标准则的流域站网优化

袁艳斌¹, 杨惜岁¹, 陈丽雅¹, 袁晓辉², 董恒¹, 俞艳¹

(1. 武汉理工大学资源与环境工程学院, 湖北 武汉 430070;
2. 华中科技大学水电与数字化工程学院, 湖北 武汉 430074)

摘要: 对于流域站网空间布局不合理(含有大量冗余站点及监测空白区等)的问题,基于美国伊利诺伊斯河下游流域 27 个站点 4 年的日均径流量数据,以站网中联合熵最大、冗余度最小且数据波动最小 3 个准则,构建了联合熵比、冗余度比和纳什效率系数(NSE)3 个目标函数对站网的信息量进行定量评估。约简流域冗余站点构成初步优化站点组合,并根据出水口得到的潜在站点确定流域需要新增的站点。结果表明:在原站网的基础上选出 11 个站点构成初步优化站点组合,其联合熵与原站网相等,冗余度减少 74.2%;在初步优化站点组合基础上新增 3 个站点,联合熵和冗余度分别增加 37.1% 和 22.3%;与原站网相比,优化后的站点数量为 14 个,联合熵增加了 37.1%,冗余度减少了 68.5%。

关键词: 水文站网;多目标优化;联合熵比;冗余度比;NSE;伊利诺伊斯河下游流域

中图分类号:TV212.5⁺1;P336 **文献标志码:**A **文章编号:**1000-1980(2019)02-0102-06

Optimization of the basin hydrologic network based on multi-objective criteria

YUAN Yanbin¹, YANG Xisui¹, CHEN Liya¹, YUAN Xiaohui², DONG Heng¹, YU Yan¹

(1. School of Resources and Environment Engineering, Wuhan University of Technology, Wuhan 430070, China;
2. College of Hydroelectric and Digitalization Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: The spatial distribution of a basin hydrological network is usually unreasonable, and the network may contain a large number of redundant stations and monitoring blank areas. In this regard, the 4-year and mean daily runoff data from 27 sites of the Lower Illinois River Basin is used in this study to achieve an optimization of hydrological network. Three objective functions, which are joint entropy ratio, redundancy ratio and Nash-Sutcliffe efficiency coefficient (NSE), are constructed to quantitatively evaluate the information of station network, based on three criteria of maximum joint entropy, minimum total correlation and minimum data fluctuation. The results show that a preliminary optimized site combination composed of 11 stations can provide the same amount of information as the original station network, and at the same time, the redundant information can be reduced by 74.2%. The total amount of information and redundant information were increased by 37.1% and 22.3% respectively by adding 3 sites on the basis of preliminary optimized site combination. Compared with the original station network, the number of optimized sites is 14, the total amount of information is increased by 37.1%, and the amount of redundant information is reduced by 68.5%.

Key words: hydrological station network; multi-objective optimization; joint entropy ratio; redundancy ratio; Nash-Sutcliffe efficiency; lower Illinois River Basin

气候变化、人类活动会改变流域水资源状况,导致水文站网的结构失调,空间布局不合理^[1-2],因此有必要对站网的布局进行重新规划。流域站网分布主要存在两方面问题:(a)流域站网存在监测空白区;(b)存

基金项目: 国家自然科学基金(41571514,51379080,41471339,41701483)
作者简介: 袁艳斌(1970—),男,教授,博士,主要从事地理信息科学研究。E-mail:yybjm@126.com
通信作者: 杨惜岁,硕士研究生。E-mail:1121695640@qq.com
引用本文: 袁艳斌,杨惜岁,陈丽雅,等. 基于多目标准则的流域站网优化[J]. 河海大学学报(自然科学版),2019,47(2):102-107.
YUAN Yanbin, YANG Xisui, CHEN Liya, et al. Optimization of the basin hydrologic network based on multi-objective criteria[J]. Journal of Hohai University(Natural Sciences),2019,47(2):102-107.

在冗余站点^[3]。为此,近年来水文站网优化研究也从对水文站网密度、现状布局以及功能评价^[4-5]方面的定性研究转为定量的优化研究^[6-8]。常用的定量研究方法有基于克里金的地统计方法^[9]、多准则决策方法^[10-11]和信息熵。基于克里金的地统计方法通过减少克里格误差优化站网,多准则决策法通过确定一个总目标及多个准则来优化站网,这 2 种方法多考虑单个站点,未能将站网作为一个整体,从而未顾及信息的可传递性、相关性和衰减性,因此在站网整体优化上存在一定的局限性。信息熵的物理机理和水流系统具有相似性^[12],因此可将信息熵应用于水文站网优化研究中。在增设站点方面,Leach 等^[13]以最大联合熵和最小总相关度为目标函数建立双熵多目标优化模型,找到流域内潜在站点的最佳位置;Su 等^[14]在互信息-距离模型的基础上提出一种改进的二维互信息-距离模型来确定潜在站点。在确定冗余站点方面,Xu 等^[15]采用基于熵理论的多准则站点优化模型,构建了最大联合熵、最小信息重叠和最小数据偏差 3 个标准,找到不同站点数目下的最优组合;Stosic 等^[16]提出基于联合排列熵的新方法应用于拉索斯河流域站网优化,使该流域布设站点数目减少了 50%。

目前,多数研究主要针对问题的一个方面对流域站网进行优化,少有对流域冗余站点和监测空白区 2 种情况开展统一研究。研究多从站网所含信息总量最大而信息冗余量最小的角度出发,采用联合熵最大、冗余度最小的双熵多目标优化函数进行站网优化,但 2 个目标函数的值域不同,因此在用参考点法求解时存在目标函数权重不一致的问题。笔者考虑对联合熵与冗余度进行归一化处理,拟定联合熵比和冗余度比 2 个目标函数,并加入 NSE 函数,从所选站网与原站网实测数据偏差来进一步判断优化站网的优劣,3 个目标函数的值域相同,当应用参考点法求解多目标优化函数时,保证了目标函数影响权重一致。因此,笔者构建联合熵比、冗余度比和 NSE 效率系数 3 个目标函数对站网进行初步优化,确定流域的冗余站点;进而通过流域分区确定流域出水口,基于初步确定的优化站网,构建多目标优化模型确定新增站点的数量和空间位置,达到“最少站点,最大联合熵且最小冗余度”的目的。

1 方 法

1.1 水文站点信息熵

在水文站网中,可把每个观测站点当作一个信息源,流量数据为信息源发出的信号,一个站点所含的信息量可用信息熵 $H(X)$ (式(1))表征^[17]。将观测流量数据序列等间隔划成 s 组,统计每组的个数, p_i 为每组的概率。

$$H(X) = \sum_{i=1}^s p_i \log_2 \frac{1}{p_i} \tag{1}$$

当站网中存在 N 个站点 $(X_1, X_2, \dots, X_N)$ 时,站点所含的信息总量可用联合熵 $H(X_1, X_2, \dots, X_N)$ 表示,如式(2)所示,其中 $p(x_1, x_2, \dots, x_N)$ 表示变量 $(X_1, X_2, \dots, X_N)$ 的联合概率密度函数。式(2)中 u 表示 x_1 的取值集合, v 表示 x_N 的取值集合。

$$H(X_1, X_2, \dots, X_N) = - \sum_{x_1 \in u} \dots \sum_{x_N \in v} p(x_1, x_2, \dots, x_N) \log_2 p(x_1, x_2, \dots, x_N) \tag{2}$$

由于站点间存在信息重叠,因此用冗余度 $C(X_1, X_2, \dots, X_N)$ 衡量站点间的信息重叠量:

$$C(X_1, X_2, \dots, X_N) = \sum_{j=1}^N H(X_j) - H(X_1, X_2, \dots, X_N) \tag{3}$$

1.2 水文站网多目标优化

由于站点间的信息具有传递性,所以站网中存在大量冗余信息,站网优化的目标是“最少站点、最大联合熵且最小冗余度”。获取不同站点数目的全部站点组合,构建站点最大联合熵比、最小冗余度比和最大纳什效率系数(Nash-Sutcliffe efficiency coefficient, NSE)这 3 个目标函数,逐步确定站网中的冗余站点和新增站点。

1.2.1 确定冗余站点

对原水文站网的 27 个站点从 1 ~ 27 进行编号,设定优化后的站点数为 d ,从 27 个站点中选取 d 个站点的全部组合 C_{27}^d , d 依次取为 1, 2, $\dots$, 27, 则不同优化站点数目下的全部站点组合为 $\sum_{d=1}^{27} C_{27}^d$ 。根据每个站点组合与原水文站网的联合熵比、冗余度比和 NSE 这 3 个准则对不同站点组合的信息量进行定量分析,选择满足这 3 个准则(式(4))的最优站点组合。

$$\begin{cases} \max F(1) = \frac{H(X_1, X_2, \dots, X_d)}{H(X_1, X_2, \dots, X_N)} & 0 \leq F(1) \leq 1 \\ \min F(2) = \frac{C(X_1, X_2, \dots, X_d)}{C(X_1, X_2, \dots, X_N)} & 0 \leq F(2) \leq 1 \\ \max F(3) = \max NSE = 1 - \frac{\sum_{t=1}^w (x_t - q_t)^2}{\sum_{t=1}^w (q_t - \bar{q})^2} & F(3) \leq 1 \\ \text{s. t.} & d = 1, 2, \dots, N \end{cases} \quad (4)$$

式中: t ——监测时段序号; w ——总的监测时段数; x_t ——站点组合在 t 时段的平均流量; q_t ——原水文站网在 t 时段的平均流量; $\bar{q}$ ——原水文站网在整个时段的平均流量; $\max F(1)$ ——包含 d 个站点的所有组合与原水文站网的联合熵比值最大值, $F(1)$ 越接近 1 表示此站点组合能够反映原水文站网的信息量越多,也越合适; $\min F(2)$ ——站点组合与原水文站网冗余度的比值最小值,值越小,表示所选站网冗余信息越小; $\max F(3)$ ——所有站点组合的 NSE 最大值,是对数据波动的检测,当 $0 \leq NSE \leq 1$ 时表示所选站点组合可信, NSE 值越接近 1 表示站点组合与原站网所测数据的偏差越小,选择的站网也就越合理。

1.2.2 确定新增站点

根据式(4)确定原水文站网中的冗余站点并剔除,剩余站点组成初步优化水文站网。但是流域中可能还存在初步优化水文站网监测不到的区域,此时需要在合适的位置增加站点提高监测效率。利用 ArcGIS 水文分析功能进行流域分区,设置子流域出口处为可能新增的潜在站点位置;并根据排水面积比和空间邻近度方法(DAR-IDW)估计潜在站点的日均径流量^[13];基于式(5)中的多目标优化确定最优潜在站点的位置和数量。

$$\begin{cases} \max F(1) = \frac{H(X_1, X_2, \dots, X_m; Y_1, Y_2, \dots, Y_k) - H(X_1, X_2, \dots, X_m)}{H(X_1, X_2, \dots, X_m)} & F(1) \geq 0 \\ \min F(2) = \frac{C(X_1, X_2, \dots, X_m; Y_1, Y_2, \dots, Y_k) - C(X_1, X_2, \dots, X_m)}{C(X_1, X_2, \dots, X_m)} & F(2) \geq 0 \\ \max F(3) = \max NSE = 1 - \frac{\sum_{t=1}^w (x_t - q_t)^2}{\sum_{t=1}^w (q_t - \bar{q})^2} & F(3) \leq 1 \\ \text{s. t.} & k = 1, 2, \dots, l \end{cases} \quad (5)$$

式中: m ——初步优化站点数; $H(X_1, X_2, \dots, X_m)$ ——初步优化站网的联合熵; $(Y_1, Y_2, \dots, Y_k)$ ——新增站点数目为 k 时的站点组合; $H(X_1, X_2, \dots, X_m; Y_1, Y_2, \dots, Y_k)$ ——在初步优化站网的基础上增加 k 个潜在站点时的新站网联合熵; $C(X_1, X_2, \dots, X_m)$ ——初步优化站网的冗余度; $C(X_1, X_2, \dots, X_m; Y_1, Y_2, \dots, Y_k)$ ——新水文站网的冗余度; $\max F(1)$ ——增设 k 个站点时,增加的联合熵与初步优化站网的联合熵最大比值; $\min F(2)$ ——增加的冗余度与初步优化站网冗余度最小比值; $\max F(3)$ ——NSE 最大值; l ——由流域出水口确定的潜在站点数目。

1.3 参考点法求解多目标优化问题

流域站网多目标优化问题有 3 个子目标 $\max F(1)$ 、 $\min F(2)$ 、 $\max F(3)$, 求出每个子目标在可行域的最优解 f_r^* , 则 $f^* = (f_1^*, f_2^*, f_3^*)$ 为理想点, 进而求出一般解到理想点的欧氏距离:

$$E = \sqrt{\sum_{r=1}^3 (f_r - f_r^*)^2} \quad (6)$$

$\min E$ 时的解即为该流域站网多目标优化问题的最优解^[6]。

2 数 据 源

以美国伊利洛伊斯河下游流域 27 个站点组成的水文站网(图 1)为研究对象,采用本文方法对该流域的水文站网布局进行优化。从 SRTM 官网获取研究区数字地形数据(DEM),利用 ArcGIS 软件对数据进行拼接裁剪、水文分析得到研究区的河网空间分布。同时,从美国地质调查局网站(USGS)获取该流域 27 个水文站

点的日均径流量数据,考虑到数据的完整性和全面性,研究年限取为 2013 年 6 月 1 日至 2017 年 6 月 1 日。

3 试验与分析

3.1 确定并删除冗余站点

分别求出不同站点数目下的任意站点组合与原站网联合熵比、冗余度比和 NSE3 个目标函数在可行域的最优解 $f^*=(1,0,1)$ 。

根据式(4)和优化站点数目 d 分别求出 3 个目标函数值到理想点的距离,并求出距离最小的最优解。不同优化站点数目下的最优解的 3 个目标函数值如图 2(a) 所示,其到理想点的距离如图 2(b)。当优化站点数为 11 时,3 个目标函数值到理想点的距离最小。因此,优化站点数为 11 时的最优解即为最优站点组合,其与原站网的联合熵比为 1,而冗余度比为 0.258,表示在原站网的基础上能够减少 74.2% 的冗余信息量,而且模型效率系数 NSE 为 0.978,与原站网的实测数据比较吻合,效率较高。

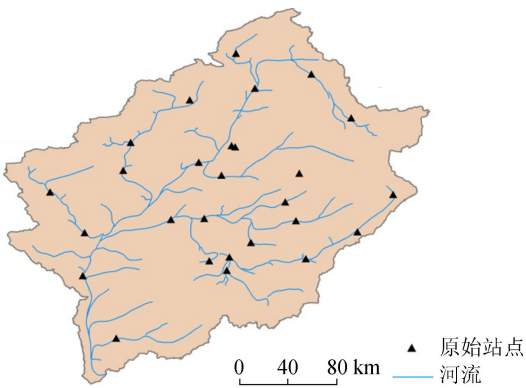


图 1 伊利洛伊斯河下游流域水文站点分布
Fig.1 Hydrological station distribution in the Lower Illinois River Basin

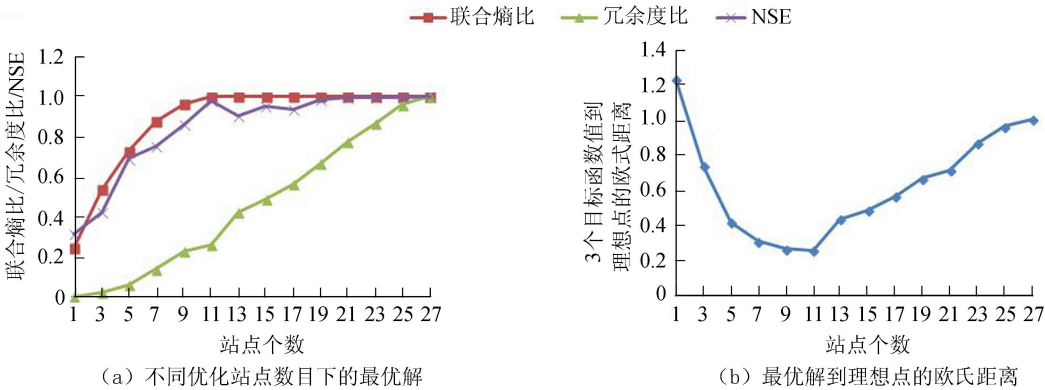


图 2 不同优化站点数目下的最优解及其到理想点的欧氏距离
Fig.2 Optimal solutions and the Euclidean distances between these solutions to ideal point with different number of optimized stations

将该最优站点组合在流域中的空间位置可视化如图 3,该站点组合即为初步优化站网。

3.2 增设潜在站点

a. 潜在站点确定:(a)对研究区的 DEM 数据进行水文分析,划分子流域和确定出水口点。(b)加入初步优化的 11 个站点,计算站点之间的距离,舍弃站点间距离小于 2 km 的其中一个站点,剩下 24 个潜在站点的位置如图4所示。

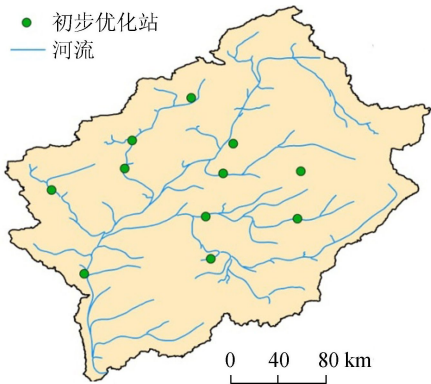


图 3 初步优化水文站网的空间分布

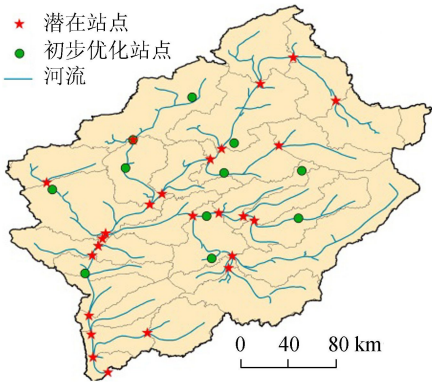


图 4 确定流域潜在站点的位置和数量
Fig.4 Determination of the location and number of potential sites in the basin

Fig.3 Spatial distribution of preliminarily optimized network

- b. 根据 DAR-IDW 方法,利用初步优化的 11 个站点实测径流量数据估计 24 个潜在站点日均径流量。
- c. 分别求出站点增加时,增加的联合熵与初步优化站网的 3 个目标函数(联合熵比、冗余度比和 NSE)在可行域的最优解即理想点。
- d. 新增站点的数目取值为 0~24,根据式(5)依次计算目标函数值,基于理想点法求出不同新增站点个数下的最优解。最优解的 3 个目标函数值变化如图 5(a)所示,其到理想点的距离如图 5(b)所示。由图 5(a)知,当新增站点数目增加到 3 时,增加的联合熵比达到最大 0.371,且不再继续增加;冗余度比在新增站点数目为 3 之前增加地比较缓慢,而之后增加的比较快;而 NSE 有下降趋势,根据 3 个目标函数具体值和趋势变化,本文只计算到增加 7 个站点即可。而由图 5(b)可知,当增加 3 个站点时,最优解到理想点的距离最小,因此该站点组合即为最优组合。该最优组合比初步优化站网的联合熵增加了 37.1%,而冗余度只增加了 22.3%,并且模型效率达到 0.98。

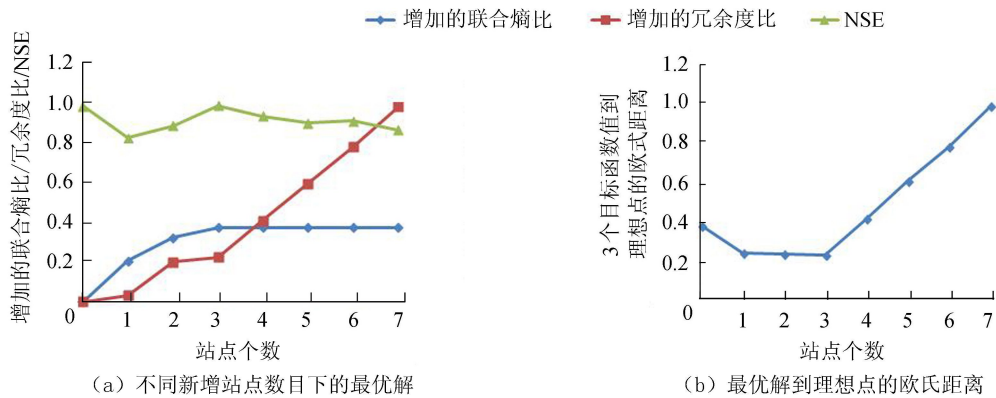


图5 不同新增站点数目下的最优解及其到理想点的欧氏距离

Fig.5 Optimal solutions and the Euclidean distance between solutions to the ideal point with different number of additional stations

- e. 根据新增站点的最优组合与初步优化站网即可确定最终的优化水文站网(图6)。

3.3 结果对比

原水文站网、初步优化水文站网以及最终优化水文站网的联合熵、冗余度和 NSE 见表 1,初步优化水文站网和原水文站网的联合熵相同,而冗余度减少了 23.855 bit,说明初步优化水文站网与原水文站网具有等量的联合熵,但减少了 74.2%的信息冗余度。最终优化站网联合熵达到了 9.770 bit,而冗余度只有 10.145 bit,比原站网的联合熵增加了 37.1%,信息冗余度减少了 68.5%。初步优化站网和最终优化站网的 NSE 分别达到 0.978 和 0.980,说明模型的效率较高。

4 结 论

- a. 将原有 27 个站点约简为 11 个站点构成初步优化站网。初步优化站网具有和原站网相等的联合熵,而冗余度减少 74.2%。
- b. 在初步优化站网的基础上,确定了 3 个新增站点并组成最终优化站网。最终优化站网的联合熵比初步优化站网增加了 37.1%,而信息冗余度增加了 22.3%。
- c. 最终优化站网由14个站点组成,与原站网相比,

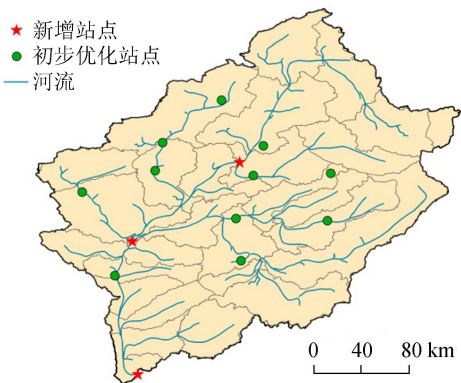


图6 最终的优化站网空间分布

Fig.6 Spatial distribution map of finally optimized station network

表1 原站网、初步优化站网和最终优化站网的联合熵、冗余度和 NSE

Table 1 Joint entropy, redundancy and NSE specific values of original hydrological network, preliminarily optimized network and finally optimized network			
站网	联合熵/bit	冗余度/bit	NSE
原站网	7.125	32.150	1.000
初步优化站网	7.125	8.295	0.978
最终优化站网	9.770	10.145	0.980

其联合熵增加了 37.1%,信息冗余度减少了 68.5%。由此可见,本文基于 3 个目标函数的流域站网优化模型,可基于较少的站点获取更多信息总量,同时能减少信息冗余量,值得推广应用于流域水文站网的评估与优化。

参考文献:

[1] 徐冉,铁强,代超,等. 雅鲁藏布江奴下水文站以上流域水文过程及其对气候变化的响应[J]. 河海大学学报(自然科学版),2015,43(4):288-293. (XU Ran, TIE Qiang, DAI Chao, et al. Study on hydrological process in upper basin of Brahmaputra River from Nuxia Hydrological Station and its response to climate change[J]. Journal of Hohai University (Natural Sciences), 2015,43(4):288-293. (in Chinese))

[2] 吴晓楷. 汤河口水文测站选址论证[J]. 水资源保护,2017,33(增刊1):19-21. (WU Xiaokai. Demonstration of location selection of Tanghekou hydrological station[J]. Water Resources Protection,2017,33(Sup1):19-21. (in Chinese))

[3] MISHRAA K, PAULIN C. Developments in hydrometric network design:a review [J]. Reviews of Geophysics, 2009,47(2):2415-2440.

[4] 何惠. 中国水文站网[J]. 水科学进展,2010,21(4):460-465. (HE Hui. China gauging station network [J]. Advances in Water Science, 2010,21(4):460-465. (in Chinese))

[5] 李海源,香天元,徐汉光. 长江流域水文站网系统评价及发展建议[J]. 人民长江,2011,42(17):20-23. (LI Haiyuan, XIANG Tianyuan, XU Hanguang. Evaluation of network system of hydrological stations of Yangtze River Basin and proposals for its development [J]. Yangtze River, 2011,42(17):20-23. (in Chinese))

[6] 陈颖,袁艳斌,袁晓辉. 基于信息熵的水文站网优化研究[J]. 水电能源科学,2013,31(7):188-190. (CHEN Ying, YUAN Yanbin, YUAN Xiaohui. Hydrologic network optimization based on information entropy[J]. Water Resource and Power,2013,31(7):188-190. (in Chinese))

[7] 李禾澍,王栋,王远坤. 基于信息熵的多目标水文站网优化探讨[J]. 南京大学学报(自然科学),2017,53(2):326-332. (LI Heshu, WANG Dong, WANG Yuankun. Entropy based multi-objective optimization for hydrologic networks [J]. Journal of Nanjing University (Nature Science),2017,53(2):326-332. (in Chinese))

[8] 张继国,吴敏,谢平,等. 基于信息熵的降雨信息区域化分析[J]. 河海大学学报(自然科学版),2013,41(6):478-481. (ZHANG Jiguo, WU Min, XIE Ping, et al. Analysis of rainfall information regionalization based on information entropy [J]. Journal of Hohai University (Natural Sciences),2013,41(6):478-481. (in Chinese))

[9] KUMAR S A, YILMAZ A G, MUTTIL N. Optimal design of rain gauge network in the Middle Yarra River catchment, Australia [J]. Hydrological Processes,2015,29:2582-2599.

[10] CEM P C, NILGUN B H. Reduction of streamflow monitoring networks by a reference point approach [J]. Journal of Hydrology,2014,512:263-273.

[11] KUMAR A K, LOHANI A K, GOEL N K, et al. Rain gauge network design for flood forecasting using multi-criteria decision analysis and clustering techniques in lower Mahanadi River Basin [J]. Journal of Hydrology,2015,4:313-332.

[12] 郭燕莎,王劲峰,殷秀兰. 地下水监测网优化方法研究综述[J]. 地理科学进展,2011,30(9):1159-1166. (GUO Yansha, WANG Jinfeng, YIN Xiulan. Review of the optimization methods for groundwater monitoring network [J]. Progress in Geography, 2011,30(9):1159-1166. (in Chinese))

[13] LEACH J M, KORNELSEN K C, SAMUEL J, et al. Hydrometric network design using streamflow signatures and indicators of hydrologic alteration [J]. Journal of Hydrology, 2015,529:1350-1359.

[14] SU H T, GENE J Y Y. Developing an entropy-based model of spatial information estimation and its application in the design of precipitation gauge networks [J]. Journal of Hydrology,2014,519:3316-3327.

[15] XU H, XU C Y, NILS R S, et al. Entropy theory based multi-criteria resampling of rain gauge networks for hydrological modeling: a case study of humid area in southern China [J]. Journal of Hydrology,2015,525:138-151.

[16] STOSIC T, STOSIC B, SINGH V P. Optimizing streamflow monitoring networks using joint permutation entropy [J]. Journal of Hydrology,2017,552:306-312.

[17] SINGH V P,张继国. 信息熵:理论与应用[M]. 北京:水利水电出版社,2012.