

吉林省河流的主成分分析

杨秉赓 王 俊 李惠明 刘长欢

翟耀宗 尚金城 纪永明

(东北师范大学, 长春)

关键词 主成分分析 河流 吉林

内 容 提 要

本文采用主成分分析方法, 通过电子计算机选取五个主成分, 对吉林省河流的水文地理空间分布, 客观地划分为十个集聚型。在分析相似性与差异性的基础上, 又用欧氏距离进行了聚类分析。

水在经济和社会发展中的作用不断提高, 是历史发展的趋势。水资源对各种问题影响的深刻化, 不仅促使人们对开发水资源的关心, 而且促进人们用新的观点与方法来重新探讨水资源分类、评价和规划。但水资源是一种动态资源, 其时空变化非常复杂, 因此必须采取先进手段进行深入分析, 才能揭示出其规律性。此文采用了现代技术, 主成分分析的数学方法, 定量的对吉林省河流进行了分类, 为重新评价吉林省水资源的分配和利用提供了科学依据, 因而它具有重大的理论意义和实际意义。

一、吉林省河流水文特征概述

吉林省位于中国东北地区中部, 在东经 $121^{\circ}38'$ — $131^{\circ}18'$, 北纬 $40^{\circ}52'$ — $46^{\circ}18'$ 之间, 全省总面积 188,486 平方公里。地势由东南向西北倾斜, 可分为东部长白山山区, 中部低山丘陵区与西部松辽平原区三部分。长白山主峰白头山, 是松花江、鸭绿江、图们江发源地。全省河长在 30 公里以上的河流有 221 条, 分属于黑龙江、辽河、图们江和鸭绿江四大水系以及绥芬河水系。江河水量的主要来源是大气降水。

吉林省水利资源丰富, 但地区分布不均衡, 年际变化大。东南部的浑江和鸭绿江中游水量丰沛, 正常年降水量达 850—1000 毫米, 正常年径流深为 400—550 毫米; 其次是头道、二道松花江, 而西部平原区为最少。江河的总水量以第二松花江水系最多, 正常年径流量为 175 亿立方米, 占全省径流量的 43%, 其次是鸭绿江和图们江流域, 分别占全省年径流量的 17% 和 14%。

水面蒸发受当地湿度、温度、风速等气象因子影响, 吉林省长白山山区气温低蒸发量小, 多年平均水面蒸发量为 800 毫米, 中部地区为 800—1100 毫米, 西部平原区气温高风速大, 水面蒸发量达 1200 毫米。

吉林省江河洪水主要是由暴雨所形成。东南部山区（鸭绿江和漫江一带）和图们江下游珲春河地区暴雨最大，多年平均最大24小时暴雨量达80—100毫米，其次是长白山天池附近约80—90毫米，暴雨量有随长白山向东和向西北方向递减趋势。而处于迎风坡的中部低山丘陵区暴雨量亦较大。

河流含沙量大小与暴雨强度、地形、土壤、植被等条件关系密切。吉林省含沙量最大的河流为东辽河，其上游地区平均年侵蚀模数达1000吨/平方公里·年，为全省之冠；其次是第二松花江支流沐石河，平均年侵蚀模数为570吨/平方公里·年；再次为小辽河、北大河、伊通河上游和辉发河上游的大沙河等河流，而以长白山山区侵蚀模数和含沙量为最小。

二、主成分分析的基本原理

主成分分析（PCA）是一种多变量数学分析方法之一。它把多个指标通过数学方法转化为少数的、新的综合指标代表原来较多的指标，而较少的综合指标既能最大限度地反映较多变量的信息，又能使较少的综合指标之间互相独立，这种处理问题的方法称为主成分分析。每个主成分都是原始变量的线性组合，经过适当的调整线性函数的系数，既可使综合指标之间互相独立，又能更集中的、更明显的代表所研究对象的主要特征。应用这些主成分能够进行河流分类和分区。主成分分析的基本思路见图1。

顺 号	计 算 流 程 及 结 果	计 算 要 点 说 明
1	原始数据输入	将水文特征值矩阵 X_i ；各元素输入计算机
2	对原始数据进行标准化 → 平均值标准差	X'_{ij} ；矩阵各元素按公式 $X_{ij} = (X_{ij} - \bar{X}_j) / \sigma_j$ 进行标准化处理
3	计算相关系数 → 得相关系数矩阵	相关系数矩阵是实对称阵，列出其三角阵即可
4	解 $ R - \lambda I = 0$ 的特征方程 → 得特征值和特征向量	用Jacobi法求解特征方程的特征根和特征向量及贡献率
5	计算因子（主成分）载荷量 → 得因子载荷矩阵	主成分载荷量反映了各主成分与原始变量间的相关程度计算公式： $\gamma^* (Z_o X_i) = \sqrt{\lambda_o} \cdot I_{oi}$
6	计算主成分得分 → 得主成分得分矩阵	根据主成分得分进行聚类分析其计算公式： $F^* = X_i^* \gamma_i^*$

图 1 主成分分析计算过程的流程图

Fig.1 Flow diagram of the computation of principal component analysis

三、吉林省河流的主成分分析

本文选取了吉林省30年以上的62个水文站，16个水文特征值共992个水文数据，进行了吉林省河流的主成分分析。

将原始水文数据输入电子计算机,按主成分分析计算流程进行计算。在计算结果的实际分析中,要注意选取主成分的标准;有的人以特征值大于1或接近于1为选取的依据;也有的人以特征值累积贡献率大于80%为选取的标准。本文选取了头五个成分作为主成分。这五个主成分代表了原始全部水文特征信息的90%左右,而损失的信息仅占10%左右,这五个主成分的特征值,特征向量和因子载荷量见表1和表2。

表 1 吉林省河流的主成分特征值及其贡献率

Table 1 Eigenvalue of principal component of the river and its contributive rate in Jilin province

主 成 分	1	2	3	4	5
特 征 值	6.96145	3.67759	1.70015	1.37013	0.65070
贡 献 率	0.435098	0.229853	0.106261	0.085635	0.040669
累积贡献率 (%)	43.51	66.5	77.12	85.68	89.75

1. 吉林省河流主成分的基本特征

第一主成分的特征值为 6.96145, 其贡献率达 43.51%, 第一主成分的因子载荷量与 X_1 、 X_7 、 X_9 、 X_{10} 、 X_{11} 、 X_{12} 、 X_{13} 有较高的正相关, 表明了第一主成分主要是代表流域的洪枯水特征。

第二主成分的特征值为 3.67759, 其贡献率达 22.99%, 它与 X_3 、 X_5 有较高的正相关, 说明第二主成分主要反映年径流的变化幅度和夏讯时期径流量的大小。

第三主成分的特征值是 1.70015, 其贡献率达 10.63%, 它仅与冰情有较高的正相关, 说明第三主成分主要是代表了河流的冰情信息。

第四主成分的特征值是 1.37013, 其贡献率为 8.56%, 它与 X_6 、 X_8 有较高的正相关, 说明第四主成分主要反映了冬季径流的特点和夏季最大平均流量的变化幅度。

第五主成分的特征值为 0.65070, 其贡献率仅为 4.1% 左右, 它突出地反映了一个流域的平均侵蚀模数的大小。

吉林省河流五个主成分的数学表达式为:

$$Z_i^* = \sum_{j=1}^{16} a_{ij} X_j^* \quad i=1,2,\dots,5 \quad j=1,2,\dots,16$$

式中, Z_i^* 为主要成分值, a_{ij} 为特征向量, X_j^* 为水文特征值。

2. 吉林省河流各主成分图象分析

按主成分分析原理,主成分可用主坐标分析图中的主轴与点群的集聚型表达(图2)。

吉林省河流的头五个主成分值综合地代表了各站水文特征信息的90%左右,而头两个主成分值的累积贡献率竟达 66.5%, 它提供了各河代表站重要水文特征信息,因此,可把对62个代表站16个水文特征的综合分析简化为对头两个主成分的分析。

现将第一主成分与第二主成分为轴的点群集聚型的水文特征加以概括,见表3。

第 I 型,反映了嫩江以大赉为代表站的水文特征,嫩江的头两个主成分得分值在全省各河中是最高的,它代表了嫩江的洪枯水径流均较大的特点。

表 2 吉林省河流的特征向量和因子载荷量
Table 2 Eigenvectors of rivers and the loadage of the factor in Jilin province

编 号	水 文 特 征 值	Z ₁ (第一主成分)			Z ₂ (第二主成分)			Z ₃ (第三主成分)			Z ₄ (第四主成分)			Z ₅ (第五主成分)		
		特征向量	因子载荷量	特征向量	特征向量	因子载荷量	特征向量	特征向量	因子载荷量	特征向量	特征向量	因子载荷量	特征向量	特征向量	因子载荷量	因子载荷量
X ₁	集 水 面 积	0.267059	0.704623	0.200284	0.384085	0.085561	0.1111562	0.185071	0.216630	-0.129591	-0.104536					
X ₂	平均年径流深	0.119559	0.315451	-0.458975	-0.880179	-0.090716	-0.118284	-0.065056	-0.076149	0.119336	0.096283					
X ₃	年径流变差系数	-0.0199828	-0.526711	0.384231	0.736841	0.055994	0.073011	0.172453	0.201861	-0.019027	-0.015348					
X ₄	3—5月多年平均径流百分比	0.083084	0.219220	-0.417880	-0.801371	0.186070	0.242616	-0.231532	-0.271015	0.268111	0.216273					
X ₅	6—9月多年平均径流百分比	-0.129515	-0.341721	0.362820	0.695781	-0.312949	-0.408053	-0.174854	-0.204671	-0.253064	-0.204136					
X ₆	10—2月多年平均径流百分比	0.168623	0.444905	-0.135825	-0.260472	0.210718	0.274755	0.627458	0.734453	0.033103	0.026703					
X ₇	最大平均流量	0.355552	0.938108	0.104795	0.200965	-0.486212	-0.063397	-0.118321	-0.138498	0.097166	0.078380					
X ₈	最大平均流量变差系数	-0.157920	-0.418665	0.246451	0.472619	0.174184	0.227118	0.429823	0.503119	0.348139	0.280829					
X ₉	最大一日洪量	0.361091	0.952723	0.121587	0.233168	-0.035334	-0.046071	-0.083124	-0.097299	0.047852	0.038600					
X ₁₀	最大三日洪量	0.361984	0.955078	0.140252	0.268961	0.022288	-0.029061	-0.052386	-0.061319	0.015300	0.012342					
X ₁₁	最大七日洪量	0.354472	0.935259	0.164882	0.316195	0.011891	0.015505	0.014353	0.061801	-0.030339	-0.024473					
X ₁₂	4—9月日平均最小流量	0.362182	0.955601	0.109370	0.209738	-0.009172	-0.011959	-0.035958	-0.042090	0.070863	0.057162					
X ₁₃	4—9月最小10天平均流量	0.362929	0.957573	0.113061	0.216818	-0.003311	-0.004343	-0.028237	-0.033053	0.061009	0.049213					
X ₁₄	多年平均侵蚀模数	-0.092134	-0.243092	0.181267	0.347617	-0.475585	-0.620115	-0.064518	-0.075520	0.760445	0.613415					
X ₁₅	最大河心冰厚	-0.044570	-0.117596	0.145290	0.278823	-0.606254	0.790493	-0.161054	-0.188517	0.332098	0.267889					
X ₁₆	冻 封 天 数	-0.103018	-0.271806	0.254547	0.488145	-0.419438	0.546905	-0.461602	-0.540317	-0.012947	-0.010443					

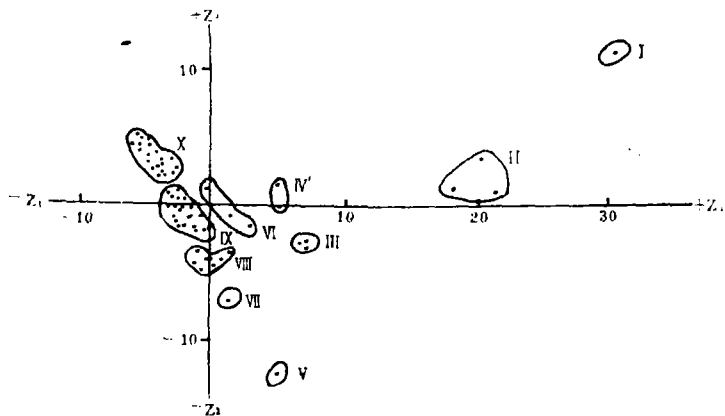


图 2 吉林省河流主成分得分图象分析
Fig.2 The image analysis of principal component scores of the rivers in Jilin province

表 3 各主成分标准得分

Table 3 The standard score of each principal component

集 聚 型	河 名	代 表 站	主 成 分 及 其 标 准 得 分				
			Z_1	Z_2	Z_3	Z_4	Z_5
I	嫩 江	大 赉	31.2932	10.3457	1.6463	2.7698	-0.5872
	鸭 绿 江	临 江	17.8679	1.2233	-1.4617	-0.4519	0.4287
II	第二松花江	红 石	21.9047	0.7268	0.0524	-1.8661	0.6543
	图 们 江	圈 河	20.7044	3.8554	-1.5731	-1.1462	-0.0041
III	头道松花江	高丽城子	7.2432	-2.8216	0.7512	-0.7542	0.3018
	二道松花江	汉 阳 屯	7.6067	-3.1197	-0.1463	-0.2734	0.0035
	浑 江	通 化	6.6889	-2.8433	-1.6128	-1.0534	0.4212
IV	辉发河	五 道 沟	5.1920	-0.0567	-0.1970	-1.5045	-0.2562
	拉 林 河	蔡 家 沟	4.8794	1.4010	0.4318	-0.8330	-0.3621
V	二道白河	二道白河	4.9876	-13.0115	-3.9020	6.5427	-0.4384
VI	布尔哈通河	么 盘 山	-0.4133	1.1350	-0.5402	-0.0355	-0.1187
	珲 春 河	桃 园 洞	1.3194	-0.7896	-0.1803	0.7427	0.0746
	图 们 河	南 坪	3.1161	-1.6599	-0.3687	0.4972	0.2066
VII	漫 江	漫 江	1.5783	-7.6355	-1.0594	-0.0297	-0.1995

第Ⅱ型，反映了第二松花江红石站和图们江的圈河站以及鸭绿江的临江站的洪枯水，径流变化幅度和夏讯时径流量都较大的水文特征。

第Ⅲ型，反映了头道松花江、二道松花江和浑江的水文特征基本相似。最大平均流量、最大平均流量变差系数、洪枯水等水文特征均相似。

第Ⅳ型，反映了辉发河和拉林河的年平均径流、洪枯水和冰情等基本相似。

第Ⅴ型，反映了二道白河及其他白河水系的独特的水文特征，其最大特点在于：第

二主成分和第四主成分具有不同于其他一切河流的年内分配的均匀性。

第Ⅶ型,反映了吉林省延吉盆地图们江上游、布尔哈通河以及珲春河的夏汛径流、洪枯水及冰情基本相似的水文特征。

第Ⅷ型,反映了发源于长白山西麓的漫江水文特征,其最大特点是年径流变差系数较小,6~9月间的多年平均径流比率稍小。

第Ⅷ型,主要反映了发源于1000米以上各中等河流水源补给相似,年径流变差系数和年内分配相近的一些河流的主要水文特征。

第Ⅸ型,包括了延吉盆地区的大部分河流和牡丹江以及发源于龙岗山脉、老爷岭、吉林哈达岭一带的绝大部分河流,其水文特征主要表现在第一主成分得分均为负值,年径流变差系数均在0.50左右,6~9月平均径流量约占年径流量的60~70%。

第Ⅹ型的河流,主要是发源于吉东低山、丘陵区和大兴安岭东麓的一些河流,其主要水文特征为:第一主成分得分均为较高的负值,第二主成分得分又为较高的正值,它突出地反映了年径流变差系数高达0.70~0.90,6~9月平均径流量约占年径流量的70~90%,东辽河流域局部地区第五主成分得分值出现极高的正值区,它恰好反映了多年平均年侵蚀模数最大区。

3. 吉林省河流主成分聚类分析

为了揭示出吉林省各条河流的类型关系,在分析了头两个主成分的相似性与差异性的基础上,又用以上五个主成分的标准平均得分距离进行了聚类分析。各类间的距离(D_{ij})以下式来表达:

$$D_{ij}^2 = \sum_{m=1}^5 (\bar{Z}_{mi} - \bar{Z}_{mj})^2$$

式中, $\bar{Z}_{mi}$ 为第*i*类(群)主成分的标准得分; $\bar{Z}_{mj}$ 为第*j*类(群)主成分的标准得分。

$m = 1, 2, \dots, 5$

经过计算得出吉林省河流的主成分标准得分谱系图3。由图中可以清楚看到,全省河流可分两大类四群十四个小型,各型之间既有一定的联系,又有差异。这些异同点,充分地反映了各流域的自然地理条件,水源补给及其各自的水文基本特征。

依五个主成分的标准平均得分距离(D_{ij})17.080可将全省河流分为两大类,即白头山北麓的二道白河和其他白河水系为一类;其余河流则为另一类。这一分类与过去的分类完全不同,它反映了二道白河与其他白河水系的自然地理条件、水源补给和水文特征与吉林省其他河流显然不同。主要差别在于:二道白河发源于白头山天池,其他白河水系皆发源于长白山自然保护区境内,该区原始林密布,地表又为火山灰及节理、熔洞较发育的玄武岩所复盖,故降水易转化为地下径流,而白河水系的绝大部分河流的水源补给以地下水为主,故这些河流年内分配较均匀,年际变化小,年径流变差系数为全东北最小区,此区水资源丰富而稳定,因此与吉林省其他河流分开。

在主成分的标准平均得分距离10.789处,将吉林省白河水系以外的其他河流又分为两群,一群为绝大多数河流均发源于以白头山为中心的多雨区,如鸭绿江的上游、图们江的上游、头道松花江和二道松花江的上游、浑江和牡丹江的上游,此区河流的水文特

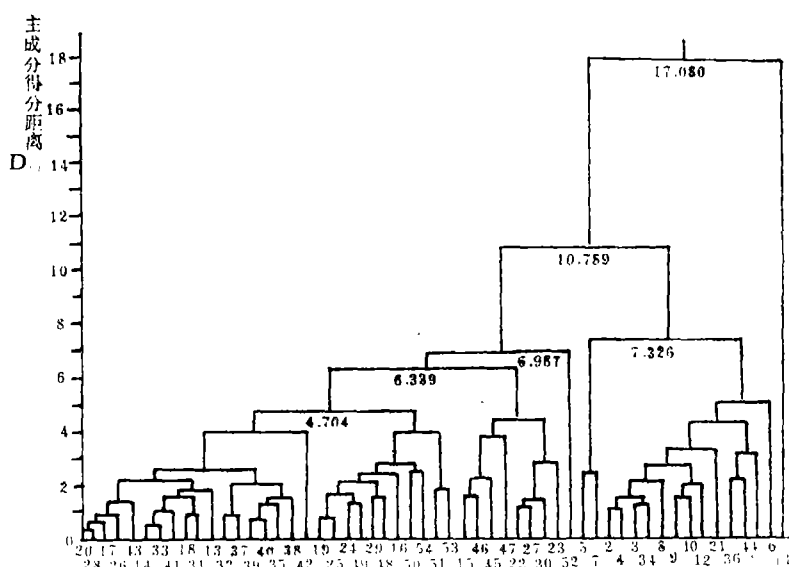


图 3 吉林省河流主成分得分聚类分析

Fig.3 The gather analysis of principal component scores of the rivers in Jilin province

1. 长白 2. 三道沟门 3. 八道江 4. 八里哨 5. 通化 6. 漫江 7. 高丽城子 8. 靖宇 9. 松江
 10. 大甸子 11. 二道白河 12. 横道子 13. 蛟河 14. 民主 15. 东丰 16. 海龙 17. 柳河
 18. 样子哨 19. 韩家 20. 口前 21. 五家子 22. 桦皮厂 23. 对龙山 24. 烟筒山 25. 官马甸
 26. 前马场 27. 刘家街 28. 榆树川 29. 伊通 30. 顺山卜 31. 七里二 32. 前红石 33. 敦化
 34. 额穆 35. 太平沟 36. 南坪 37. 团结 38. 庙岭 39. 大仙 40. 劳动 41. 邱山 42. 凤阳
 43. 英城 44. 桃园洞 45. 河清 46. 河信 47. 振德 48. 周户屯 49. 梨树 50. 索伦 51. 大石寨
 52. 保隆 53. 高家屯 54. 务本

征是：3～5 月经流量占年径流量的 20～35%，6～9 月经流量占年径流量的 55～65%，10～2 月经流量占年径流量的 10～15%，年径流变差系数在 0.20～0.40。该区水量丰富为山区灌溉和水力发电提供了良好条件。另一群则为发源于龙岗山脉以西和哈尔滨以东以及大兴安岭东麓的一些河流。这些河流的水文特征：3～5 月经流量占年径流量的 10～25%，6～9 月经流量占年径流量的 60～90%，10～2 月经流量占年径流量的 5～10% 左右，年径流变差系数达 0.50～0.90。由此可见，这些河流年内分配不均，而径流的年际变化又最大，因此与前一群分开。

在发源于以白头山为中心的河流中，主成分的标准平均得分距离为 7.326 处又分为两种类型，一是远离白头山的浑江通化站与头道松花江的高丽城子集聚在一起；另一类型是以直接发源于白头山的一些河流集聚在一起，在这一小类中虽然有的河流不是直接发源于白头山，但在一些水文特征值上它们是很相似的，所以它们也就划归这一小类了。前一类与后一类的主要差别在于：漫江和头道松花江的最大平均流量、最大一日、三日和七日洪量均值和 6～9 月平均最小流量均较后一小类大 5～10 倍。

在主成分的标准平均得分距离 6.987 处，将发源于大兴安岭东麓和龙岗山脉西侧以及哈尔滨以东的诸河流又分为五个类型。在这五类中有的又可分出次一级的类型，例如：在主成分的标准平均得分距离 4.423 处又可分出两个小型，一为发源于吉林哈达岭

南端的一些河流分在一起；一为发源于吉林哈达岭北端及土们岭以西的一些河流则又另分为一小型。它们之间的差别在于：发源于吉林哈达岭南端的一些河流，3～5月径流量占年径流量的10%以上，而发源于哈达岭北端的一些河流则春季径流不到10%；而6～9月径流后者较前者更为集中，达90%以上；在多年平均侵蚀模数方面，前者较后者大4倍以上，为全省水土流失最严重区。

综上所述，应用主成分分析法对吉林省河流进行了主成分特征值图象分析和聚类分析，它能综合性地、较完善地反映出吉林省各河流的基本特征，它是符合吉林省河流的实际情况的。这种方法对研究其他地理问题也是有效的。

参 考 文 献

- 〔1〕于崇文等：《数学地质的方法与应用》，冶金工业出版社，1980。
- 〔2〕刘承作、孙惠文：《数学地质基本方法及应用》，地质出版社，1981。
- 〔3〕吉林省水文总站、吉林水利勘测设计院：《吉林省水文图集》，1974。

THE PRINCIPAL COMPONENT ANALYSIS OF THE RIVER IN JILIN PROVINCE

Yang Binggeng Liu Changhuan Zai Yaozong Li Huiming

Wang Jun Shang Jincheng Ji Yongming

(North-eastern Normal University)

Key words: Principal component analysis; River; Jilin province

ABSTRACT

By means of principal component analysis, based on 5 principal components chosen by computer, the author divides the spatial distributions of river into gathering models. This paper also proceeds the gathering analysis using Euclidean distance, which is based on the analysis of the similarities and differences of the spatial distributions.