

澜沧江干流径流数据处理方法的应用研究

袁希平¹, 何大明²

(1. 昆明理工大学 成人教育学院, 云南 昆明 650093; 2. 云南大学 亚洲国际河流中心, 云南 昆明 650091)

摘要: 河道径流是流域水资源中最重要的组成部分. 回归相关分析作为研究随机现象的重要数学方法, 在水资源不确定性分析中具有广泛应用. 基于对澜沧江干流三个典型站点收集到的径流资料, 通过相关回归分析应用, 结果表明: 澜沧江上、中、下游三个典型站点的径流资料存在显著回归相关性, 对收集到的不完善径流数据, 可以进行适当插补和内延与外展应用, 满足了研究中
对典型站点径流数据统一性的需要.

关键词: 澜沧江; 数据处理; 回归分析; 河道径流

中图分类号: X832 **文献标识码:** A **文章编号:** 1007 - 855X(2004)03 - 0001 - 04

Application Regression Analysis on the Data Processing in Lancang River Runoff

YUAN Xi-ping¹, HE Da-ming²

(1. Faculty of Continuing Education, Kunming University of Science and Technology, Kunming 650093, China;

2. Asian International Rivers Centre, Yunnan University, Kunming 650091, China)

Abstract: The river runoff is the main part of basin water resources. As one of the mathematical ways to study the random phenomena, regression analysis is the important tool to study the uncertainty of basin water resources. Based on the river runoff materials collected from three typical different survey stations located spatially in upper, middle and down stream of Lancang River, regression analysis is made, and the result is that there exists beeline regression relativity in the materials from the three different survey stations, so it is possible to make a loss for the runoff materials in these survey stations.

Key words: Lancang River; data processing; regression analysis; river channel runoff

0 引言

跨境资源与环境问题是目前世界极为关注的热点之一. 国际河流流域的水资源, 打破了国家领土的界线, “自由”地从一个国家流向另一个国家, 为流域各国所共享. 澜沧江—湄公河源于我国青海省, 流经中国、缅甸、老挝、泰国、柬埔寨和越南 6 国, 战略位置极为重要. 针对澜沧江流域跨境水资源展开有关研究, 分析掌握水资源径流状况与变化趋势, 对该流域的地区合作与发展具有重要的社会现实意义.

河道径流是流域水资源中最重要的组成部分. 受多种自然因素的影响, 河道径流水文现象具有不确定性, 但对于长时间的系列观测结果, 则存在随机不确定性意义上的统计规律. 分析研究河流年径流系列特征与统计规律, 对于认识流域水资源状况, 指导水资源开发利用、控制洪涝灾害、促进水资源的合理分配等具有十分重要的基础作用.

数理统计分析作为研究随机现象的重要数学方法, 在水资源不确定性研究中具有广泛的应用. 根据高斯-马尔可夫定理的基本结论, 最小二乘估计法作为求算回归方程式中参数的最一般方法, 由其计算而得的估计量在所有的线性无偏估计中具有最小方差性. 依据该结论所进行的直线相关分析, 是研究两个变量之间存在的随机不确定性关系最常用的方法. 文中针对澜沧江流域水资源分配利用研究, 基于对澜沧江干流三个典

收稿日期: 2004 - 03 - 01. 基金项目: 国家“十五”科技攻关西部重大项目(项目编号: 2002BA901A22); 国家重点基础研究发展计划专题(项目编号: 2003CB415105).

第一作者简介: 袁希平(1965. 1 ~), 男, 副教授. 主要研究方向: 自然资源管理与 3S 技术应用. E-mail: YXPCS@263.net

型站点收集得到的多年不同起点和末点的年径流数据,通过相关分析,利用科学处理,对缺少或不足资料进行插补或延长,以满足进一步研究中对三个典型站点年径流资料数据统一性与完整性的要求。

1 研究方法

回归相关分析是研究变量之间关系,建立变量之间回归方程模型,判断变量之间关系密切程度的统计方法,仅考虑一个自变量时的回归相关分析被称为一元回归分析。一元回归分析用于研究两个变量之间是否存在直线关系,如何建立回归方程,探讨两变量之间统计关系,判断相关程度,最后进行插补和外延。

1.1 一元回归相关分析方法的数学描述与参量估算

若对于变量 x 的一组不全相等的值 (x_1, x_2, \dots, x_n) 作独立试验,得到随机变量 y 的相应观测值 (y_1, y_2, \dots, y_n) ,如此构成的对数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 称为一组大小为 n 的样本。对于 x 的每一个值,若 y 服从正态分布 $N(ax + b, \sigma^2)$,则 $\tilde{y} = a + bx$ 将称为理论回归方程, b 称为回归系数。其中 a, b, σ^2 是参量常数。

在回归分析中,一般用样本 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 来估计 a, b ,从而得到估计参数值 \hat{a}, \hat{b} ,由此得到理论回归方程 $\tilde{y} = a + bx$ 的一个估计: $\hat{y} = \hat{a} + \hat{b}x$,即称为回归方程。

根据最小二乘估计方法,得到 \hat{a}, \hat{b} 估计,并定义相关系数 r 的计算公式如下:

$$b = \frac{(x_i - \bar{x})(y_i - \bar{y})}{(x_i - \bar{x})^2}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

$$r = \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(x_i - \bar{x})^2} \sqrt{(y_i - \bar{y})^2}}$$

1.2 线性回归显著性检验及相关分析

从统计学角度看,回归和相关分析是有区别的。一般地,回归反映的是两变量间存在的依存关系,而相关则反映两变量之间所存在关系的密切程度等,但在实际应用中,常将两者结合在一起应用,如通过绘制散点图,直观判识相关趋势,选择回归模型,进行最小二乘拟合估算得到回归方程,计算相关系数并同时同时进行回归检验与相关分析,在回归相关分析达到显著水平之后,再利用回归方程进行插补和外延预测等。

线性回归相关分析显著性检验通常用 F 检验法,即在原假设 $H_0: b = 0$ 条件下,若 $F = (n - 2) \frac{r^2}{1 - r^2}$ 大于分布值 $F_{1 - \alpha}(1, n)$,则以 $(1 - \alpha)$ 的置信水平拒绝原假设,亦即以 $(1 - \alpha)$ 置信水平相信备假设,反之则接受原假设,拒绝备假设。

2 方法应用研究

2.1 资料收集与整理

在澜沧江干流多个水文测站中,考虑站点空间分布的代表性、资料积累的完备性、以及对流域上、中、下游径流的控制性作用等,选取了上游的昌都、中游的旧州、下游的景洪3个水文站点作为有关径流分析的研究对象。3个站点在澜沧江干流的空间分布示意如图1所示。

对3个站点的有关多年径流资料进行收集、整理,并依据有关计算公式处理得到3个典型站点多年径流均值、变差系数(C_v)及偏态系数(C_s)等,3个站点收集资料基本信息及有关处理结果如表1所列。相应多年年径流系列过程动态绘制如图2所示。



图1 3个站点在澜沧江干流的空间分布示意

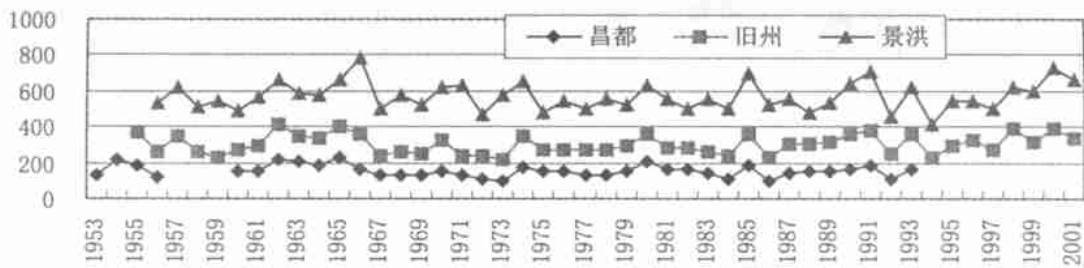


图2 3个典型站点年径流收集资料整理系列过程示意

2.2 参数计算与回归相关分析

从以上整理结果可看出,对澜沧江干流典型站点收集到的资料,起止时间系列不全一致,另外,昌都站还有几个中间年份的数据不完整.为了开展澜沧江干流相关径流分析研究,需要将其进行必要的技术处理,以获得一致的时间系列,为开展流域水资源研究提供统一的信息工作基础.为此,首先需要对3个站点之间的有关收集资料,依据上述统计原理方法,进行回归相关分析,然后对收集到的时间系列不一致的数据进行必要的插补和内延外展处理.

依据统计分析原理方法,基于收集资料,分别计算上游昌都与中游旧州、中游旧州与下游景洪站点、上游昌都与下游景洪站点之间径流资料的相关统计特征,获得回归方程,并进行相应的相关性分析.具体处理过程中,由于在1960~1993共34年间,3个站点都具有完善的系列径流资料,因此提取该时段系列径流资料,进行两站点之间的年径流资料回归相关分析.

表1 典型站点资料收集基本信息及其整理

流域位置	站点名	时间序列	年径流/ 亿 m ³ a ⁻¹	方差	多年变差 系数/ C _v	偏态系 数/ C _s	最大最小 比值
上游	昌都站	1953~1993	153	32.46	0.213	0.34	2.4347
中游	旧州站	1955~2001	302	52.78	0.176	0.44	1.8521
下游	景洪站	1956~2001	570	78.03	0.137	0.74	1.8887

依据收集资料分别绘制两站点间相互的径流散点图3~图5.观察散点图分布状况可直观看出,任意两站点之间均具有直线分布态势,为此采用直线模型,按统计原理进行拟合估算,获得不同两站点之间的回归系数、回归方程、相关系数等如表2所列.

表2 两站点间的回归方程相关性检验结果整理

	回归直线方程	相关系数 平方	相关 系数	F 计算值	相关 效果
昌都与旧州	$y = 1.4631x + 76.96$	0.8099	0.90	136.35	显著
昌都与景洪	$y = 1.4092x + 356.11$	0.3437	0.74	39.55	显著
旧州与景洪	$y = 1.0991x + 241.15$	0.5527	0.59	16.76	显著

首先由回归系数符号分析可得知,昌都站与旧州站、旧州站与景洪站之间的径流均存在直线正相关关系,即中游旧州站、下游景洪站径流均具有随上游昌都站径流增加而直线增加趋势,下游景洪站径流也具有随中游旧州站径流增加而直线增加趋势,但就回归系数绝对值看,相对与下游,中游旧州站径流随上游昌都站径流增加而增加的趋势更为明显.

其次,为了对获得的直线方程回归相关性进行分析,采用F检验法对3条直线回归相关效果分别进行F检验:首先以99%为置信水平,基于采用的样本数为34年,在F统计分布表中查得 $F_{1, (1, N - 2)} = F_{0.99}(1, 32) = 7.50$,依据计算得到的3条拟合直线相关系数(0.90, 0.74, 0.59),按公式计算得到相应的F值(164.32, 39.55, 16.76),参照对比 $F_{0.99}(1, 32) = 7.50$,结果得出,由于3条回归拟合直线的F值(164.32, 39.55, 16.76)均大于7.50,因此按统计推理得出,3直线均以99%的置信水平拒绝原假设(回归直线不相关),说明上游昌都与中游旧州、中游旧州与下游景洪、上游昌都与下游景洪的径流资料都具有显著的回归相关效果.换言之,基于本分析结果可得出,在澜沧江流域有关径流研究中,可以在较高的置信度水平下(99%),利用已知站点的观测资料对其它站点所缺乏资料进行适当插补和内延与外展处理.

2.3 插补和内延与外展处理

基于以上回归及相关分析,对3个典型站点资料进行必要的数据统一与完善技术处理,即利用已验证

相关回归效果显著的回归方程对收集资料进行必要的插补和内延与外展处理.具体地,利用回归方程 1 对上游昌都站点的 1957~1963 年所缺数据进行插补处理,对 1994~2001 年所缺数据进行外展处理,对中游 1953 年和 1954 年进行内延处理;利用回归方程 2 对下游 1953~1955 年资料进行内延处理.处理后 3 个站点的系列过程与回归系列过程 6 所示.

3 结语

在澜沧江径流数据处理中,通过回归相关分析应用结果表明:(1)上游昌都与中游旧州、中游旧州与下游景洪、上游昌都与下游景洪的径流资料都存在显著的回归相关性,因此,对于干流上的这 3 个站点收集

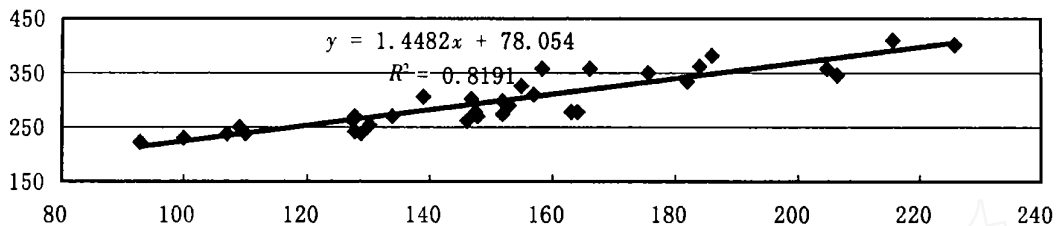


图3 上游昌都与中游旧州站点间径流散点图与直线拟合图

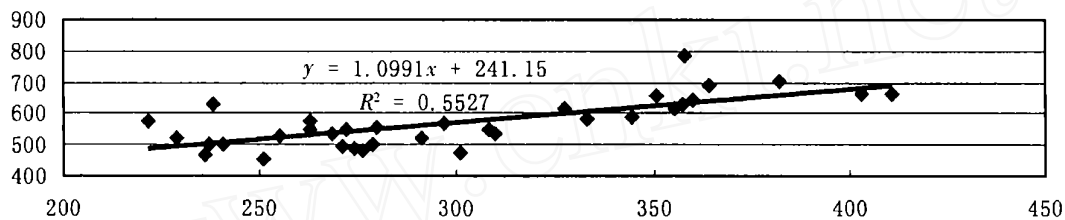


图4 中游旧州与下游景洪站点间径流散点图与直线拟合图

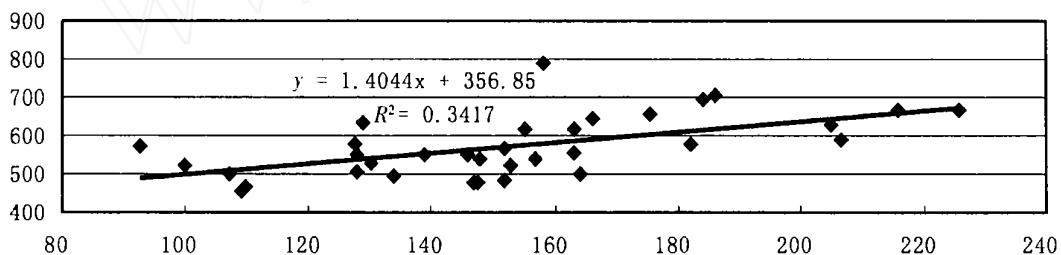


图5 上游昌都与下游景洪站点间径流散点图与直线拟合图

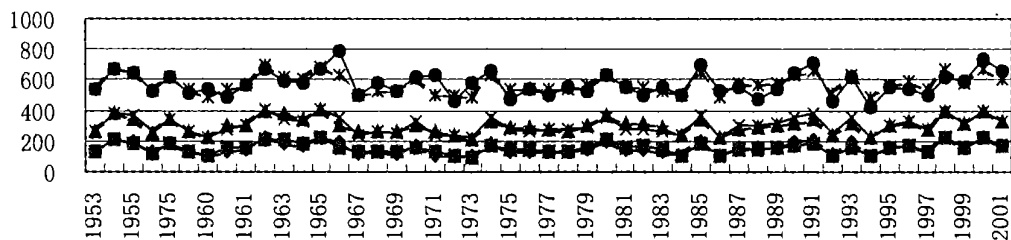


图6 3个典型站点年径流收集处理后整理系列过程示意

到的不完善水文径流系列数据,可以在较高的置信度水平下进行适当插补和内延与外展处理;(2)回归相关分析的目的在于探索两个变量间存在的相关模式与相关程度等问题,由此应用于专业分析时,需要特别注意的是,回归相关分析只是一种数学工具,对于应用中的有关规律认识与分析需要密切结合专业理论知识,以一定的科学依据为前提,只有这样才能有效的应用于水文分析与其它工程应用当中.

(下转第12页)

第一步在 MAPGIS 的图像编辑模块中,将遥感影像文件转化为 MAPGIS 兼容的格式(.msi 或.tif).

第二步以三江地区地理底图为基准,在 MAPGIS 镶嵌配准模块中,对遥感图像进行配准,即统一两者的地理坐标.

第三步在 MAPGIS 的输入编辑功能中,以配准后的遥感图像作为工作底图,叠加融合相关的专业要素,如点(如地名、文字说明等)、线(如旅游线路、交通线等)、区(如三江范围等).

第四步整饰成图(图6).

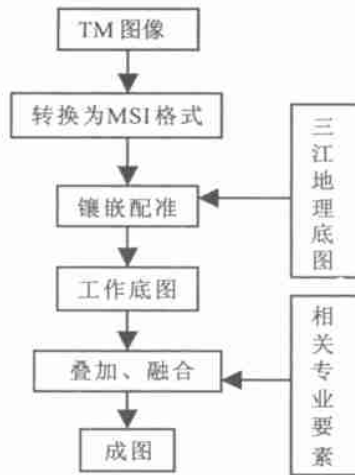


图5 三江项目遥感影像图制作程序 图6 三江并流带旅游地质精品线路图

3 问题与探讨

三江项目制图工作量大,任务紧,图件主要由插图、图版和专题图组成。项目组成员采用 word ,excel ,photoshop ,MAPGIS 等软件,以统一的流程标准要求、灵活的形式,较好地完成任务,但在一些方面还存在不足,值得进一步探索。一方面是三江地区立体效果图由于工作量大、时间紧,未能完成,成为一大遗憾,借助于数字高程模型或遥感影像加强该方面的探索极有必要;另一方面是制图人员较多,时间紧,图式图面略显杂乱,此外,图件创新仍需进一步加强。

参考文献:

- [1] 杨世瑜,王树芬,等.三江并流带旅游地质资源开发与环境保护[M].昆明:云南民族出版社,2003.1~9.
- [2] 王晟.Word2000 中文版使用方法与技巧[M].北京:人民邮电出版社,1999.137~165.
- [3] 晨禾工作室.Excel2000 中文版实用问题解答[M].北京:人民邮电出版社,1999.193~220.
- [4] 沈大林.photoshop7.0 中文版案例教程[M].北京:中国铁道出版社,2003.1~84.

(上接第4页)

参考文献:

- [1] 何大明.澜沧江-湄公河水文特征分析[J].云南地理环境研究,1995,7(1):58~74.
- [2] 李丽娟,李海滨,王娟.澜沧江水文与水环境特征及其时空分异[J].地理科学,2002,22(1):49~56.
- [3] 何大明.通过水资源整体多目标利用和管理推进澜沧江-湄公河流域的持续发展[J].云南地理环境研究,1996,8(1):25~36.
- [4] 甘淑,王人潮,何大明.澜沧江流域云南段土地覆盖状况及其模式分类[J].山地学报,2000,18(5):422~428.
- [5] 陈丽晖,何大明.澜沧江——湄公河整体水分配[J].经济地理,2001,21(1):28~32.
- [6] 刘恒,刘九夫,唐海行.澜沧江流域(云南段)水资源开发利用现状及趋势分析[J].水科学进展,1998,119(11):70~76.
- [7] 李秀云,傅肃性,李丽娟.河流枯水极值分析与模型预测研究[J].资源科学,2000,22(5):73~77.
- [8] 袁希平,何大明.澜沧江流域典型案例区地形地貌对比分析[J].昆明理工大学学报(理工版),2002,27,(6):1~5.