

车尔臣河春季来水分类概率预报

苏 仁¹ 琪美格²

1. 新疆巴音郭楞水文勘测中心 新疆 库尔勒 841000

2. 新疆维吾尔自治区水文分析计算中心 新疆 乌鲁木齐 830000

摘 要: 车尔臣河发源于昆仑山北坡, 位于新疆且末县, 且末水文站是其水量控制站。论文选用该站1958-2021年春季来水(4至5月平均流量)为预报对象, 并构建为由来水偏丰、正常或偏枯组成的分类因变量, 再由逐步回归分析技术优选所得的3月月平均流量、前一年9月月平均流量、前一年12月月降水量、2月月平均气温为预报因子, 采用多项Logistic回归分析技术建立了回归方程, 对2022年春季来水每类可能发生的概率进行了定性预报, 结果正确。

关键词: 车尔臣河; 分类因变量; 多项Logistic回归分析; 概率预报

引言

车尔臣河径流机制是以冰雪融水为主、降水补给次之, 受春季升温的影响, 历史上年最大洪峰流量曾发生在春季(4至5月), 导致该河4至5月来水定量预报难度增大, 所以有必要在定量预报之前通过定性预报来把握其演变趋势, 防止定量预报发生大偏差。论文通过优选影响且末水文站春季来水的预报因子, 将春季来水构建为由偏丰、正常或偏枯组成的分类因变量, 采用多项Logistic回归分析技术, 对分类因变量每类可能发生的概率进行定性预报, 从而有效把握春季来水未来演变趋势, 对避免定量预报发生大偏差并确保预报精度有重要意义。

1 技术思路与方法

1.1 预报因子优选技术

春季来水情况对于水资源规划、农业灌溉等至关重要, 准确预报春季来水离不开科学选取预报因子。逐步回归分析为解决这一问题提供了有效途径。该方法采用引进和剔除因子的双重F检验机制。在分析过程中, 先从众多潜在预报因子中, 依据F检验筛选出对预报对象(春季来水)影响显著的因子引入回归方程; 同时, 持续监测已引入方程的因子, 若某个因子对预报对象的影响不再显著, 便将其剔除。通过这种动态调整, 不断优化回归方程。借助逐步回归分析, 能从海量预报因子中精准优选出具有统计学意义的主要因子, 构建出科学合理的线性回归模型, 清晰呈现春季来水与各主要因子间的线性相关关系, 进而提高春季来水预报的准确性和可靠性, 为相关领域决策提供有力依据。

1.2 分类概率预报方法

依据水文情报预报规范, 时段流量距平是判断流量状况关键指标, 据此将4至5月平均流量设为含偏丰、正

常、偏枯三类的分类因变量。为科学定性预报这三类发生概率, 采用多项Logistic回归分析技术, 该技术适用于多分类因变量, 能剖析预报因子与因变量关系^[1]。实际操作中收集前期降水量等预报因子数据, 纳入模型运算, 确定各因子对不同流量类别的影响, 得出发生概率。借此可提前定性判断流量状况, 为相关工作提供决策依据, 降低风险损失。

多项Logistic回归分析是用数个二值Logistic回归方程, 来确定分类因变量中每类在与参照类相比时所优选的预报因子对预报对象的贡献。

若分类后的预报对象有M个类别, 取 $m = 1, 2, \dots, M$, 设第m类的概率为 G_m , 则预报对象实测值在M个类别中的分布服从多项分布, $\sum G_m = 1$ 。取 $i = 1, 2, \dots, n$ (n 为预报因子总数), 用 Z_i 表示预报因子, 用 c_m 和 d_{mi} 表示第m类的常数和预报因子回归系数, 则第m类的多项Logistic回归方程为:

$$\ln(G_m/G_M) = c_m + d_{m1} * Z_1 + \dots + d_{mi} * Z_i + \dots + d_{mn} * Z_n$$

以上回归方程的参照类是分类因变量中的最后一类M, 说明每个m类与参照类M之间建立了M-1个二值Logistic回归方程。设 $G = \sum (\text{Exp}(c_m + d_{m1} * Z_1 + \dots + d_{mi} * Z_i + \dots + d_{mn} * Z_n))$, 则分类因变量第m类可能发生的概率 G_m 为:

$$G_m = \text{Exp}(c_m + d_{m1} * Z_1 + \dots + d_{mi} * Z_i + \dots + d_{mn} * Z_n) / G$$

由于参照类M相应的常数与回归系数为0, 所以参照类M可能发生的概率 G_M 为1/G。

上述M个概率中的最大值对应的类别就是最可能发生的类别, 即为概率预报结果。

2 优选影响春季来水的预报因子

在充分利用现有水文气象资料条件下, 选用且末水文站1958-2021年4至5月月平均流量为预报对象, 1-3月月平均流量及前1年1-12月月平均流量、月平均气温、月降

水量等45个序列为预报因子，用逐步回归分析技术优选了影响预报对象的通过信度为0.05双重F检验的4个预报因子，按影响程度从大到小排序，依序是3月月平均流量、前一年9月月平均流量、前一年12月月降水量和2月月平均气温，分别用Z1、Z2、Z3和Z4来表示。

经统计，逐步回归方程复相关系数为0.772，表明回归方程代表性好；统计量 $F = 21.7 > F(0.001) = 5.32$ ，表明预报因子与预报对象之间有线性回归关系，优选结果有统计学意义。

3 春季来水分类与概率预报

春季来水的分类与概率预报，需要应用SPSS数据文件。且末水文站1958-2021年3月月平均流量Z1、前一年9月月平均流量Z2、前一年12月月降水量Z3、2月月平均气温Z4及4至5月平均流量Q序列SPSS数据文件见图1、图2。

3.1 构建春季来水分类因变量

经统计，4至5月平均流量多年均值为 $22.9\text{m}^3/\text{s}$ ，其正、负10距平是 $25.2\text{m}^3/\text{s}$ 和 $20.6\text{m}^3/\text{s}$ 。用W表示分类因变量，则据前文规范规定，Q小于 $20.6\text{m}^3/\text{s}$ 时，来水偏枯，对应的分类因变量W取1；Q在 $20.6\text{m}^3/\text{s}$ 和 $25.2\text{m}^3/\text{s}$ 之间为来水正常，W取2；大于 $25.2\text{m}^3/\text{s}$ 为来水偏丰，W取3。由此可构建4至5月平均流量分类因变量W序列，见图1、图2。

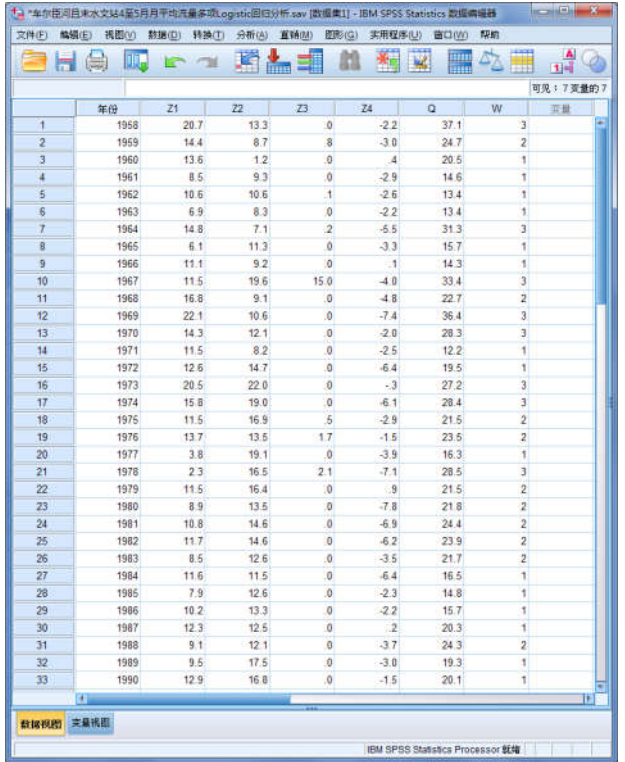


图1 且末水文站春季来水分类因变量及预报因子序列 (一)

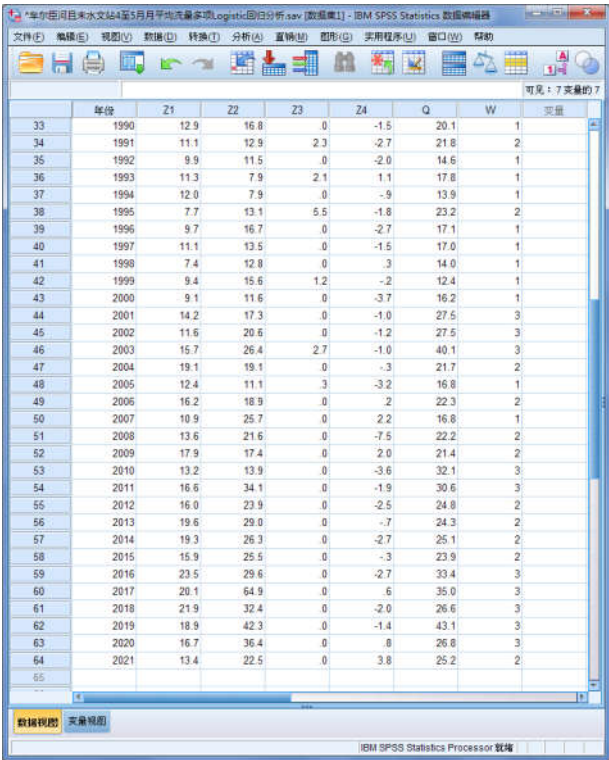


图2 且末水文站春季来水分类因变量及预报因子序列 (二)

3.2 多项Logistic回归分析

可由SPSS统计软件来实现多项Logistic回归分析，SPSS实现过程为：

(1) 依次单击图1中的菜单“分析→回归→多项Logistic”，在弹出的多项Logistic回归对话框中，将W和Z1、Z2、Z3、Z4分别移至因变量和协变量列表框。

(2) 完成变量设置后，要获取更深入分析结果，需细致设置统计量。单击“统计量”按钮，在弹出对话框中，除默认勾选的基础统计项外，增选“分类表”意义重大，它能以表格直观呈现模型对不同类别的预测情况，如准确率等。增选“拟合度”可评估模型与实际数据的契合度，辅助优化模型。设置完单击“继续”，软件保存设置并返回主界面。

(3) 各项设置完备，单击“确定”按钮开启分析。软件会依据设定，运用算法处理数据，自动生成回归效果的统计检验结果。这些结果涵盖模型关键信息，是解读数据、辅助决策的重要依据。

3.3 回归效果的统计检验

回归效果的统计检验有6个主要结果：

(1) 案例处理摘要：针对分类因变量进行了细致梳理。结果显示，偏枯类别的个案数量为25个，呈现出水资源相对匮乏的状态；正常类别的个案有21个，表明水

资源状况处于常规水平；偏丰类别的个案达18个，意味着水资源较为充足。三类个案合计64个，且所有个案均有效，为后续准确分析提供了坚实的数据基础。

(2) 回归方程拟合信息：用-2倍对数似然值作为回归方程拟合标准，对仅含截距项和最终的两个回归方程进行似然比检验，结果表明，其-2倍对数似然值分别为139.470、83.778，卡方为55.692，自由度为8，显著性水平为0.000小于0.001，说明最终回归方程明显优于仅含截距项的回归方程，且最终回归方程中至少有1个预报因子有统计学意义。

(3) 拟合优度检验：Pearson卡方为91.305，显著性水平为0.967大于0.05，而偏差卡方为83.778，显著性水平为0.993也大于0.05，说明回归方程预测值与实测值的差异无统计学意义，拟合优度较好^[3]。

(4) 似然比检验：优选的4个预报因子的显著性水平都小于0.005，说明4个预报因子对回归方程的贡献都有统计学意义。

(5) 回归方程参数估计：分类因变量W每类可能发生的预测概率 G_m （ m 取1、2、3）计算式分别为：

$$G_1 = \frac{\text{Exp}(16.263 - 0.677 \cdot Z_1 - 0.347 \cdot Z_2 - 1.728 \cdot Z_3 + 0.644 \cdot Z_4)}{(1 + \text{Exp}(16.263 - 0.677 \cdot Z_1 - 0.347 \cdot Z_2 - 1.728 \cdot Z_3 + 0.644 \cdot Z_4) + \text{Exp}(4.828 - 0.157 \cdot Z_1 - 0.079 \cdot Z_2 - 0.265 \cdot Z_3 + 0.211 \cdot Z_4))}$$

$$G_2 = \frac{\text{Exp}(4.828 - 0.157 \cdot Z_1 - 0.079 \cdot Z_2 - 0.265 \cdot Z_3 + 0.211 \cdot Z_4)}{(1 + \text{Exp}(16.263 - 0.677 \cdot Z_1 - 0.347 \cdot Z_2 - 1.728 \cdot Z_3 + 0.644 \cdot Z_4) + \text{Exp}(4.828 - 0.157 \cdot Z_1 - 0.079 \cdot Z_2 - 0.265 \cdot Z_3 + 0.211 \cdot Z_4))}$$

$$G_3 = \frac{1}{(1 + \text{Exp}(16.263 - 0.677 \cdot Z_1 - 0.347 \cdot Z_2 - 1.728 \cdot Z_3 + 0.644 \cdot Z_4) + \text{Exp}(4.828 - 0.157 \cdot Z_1 - 0.079 \cdot Z_2 - 0.265 \cdot Z_3 + 0.211 \cdot Z_4))}$$

上述计算式中， G_1 是类别1与参照类3相比计得的来水偏枯预测概率， G_2 是类别2与参照类3相比计得的来水正常预测概率， G_3 是参照类3的来水偏丰预测概率。

(6) 分类表中预测正确的总百分比达70.3%，说明回归方程预测效果很好。

3.4 2022年春季来水概率预报

且末水文站2022年3月月平均流量 Z_1 为19.3m³/s，前一年9月月平均流量 Z_2 为17.5m³/s，前一年12月月降水量 Z_3 为0.0mm，2月月平均气温 Z_4 为-1.2℃，代入上述3个不同类别的预测概率计算式，得到 G_1 、 G_2 、 G_3 分别为0.012、0.534和0.454，其中， G_2 值最大，说明分类因变量W为2的可能性最大，即且末水文站2022年4至5月平均流量为正常的可能性最大，实况来水是23.4m³/s，介于正常的上限值25.2m³/s和下限值20.6m³/s之间，春季来水属于正常，定性预报正确。

结语

(1) 基于逐步回归优选预报因子，用多项Logistic回归分析对春季来水分类因变量概率定性预报，是中长期水文预报的成功拓展，对避免大偏差、确保精度意义重大。

(2) 多项Logistic回归方程拟合信息佳，最终方程至少1个因子有统计学意义，拟合优度好，优选因子贡献显著，预测正确率达70.3%。

(3) 2022年春季来水概率预报结果正确，证明该方法适用于中长期水文定性预报。

参考文献

- [1] 琪美格. 白杨河4至5月平均流量多项Logistic回归分析与预报[J]. 地下水, 2021(04): 203-204.
- [2] 苏仁, 旦木仁加甫. 新疆车尔臣河4至5月来水成因分析与预报[J]. 河南水利与南水北调, 2023(02): 40-44.
- [3] 苏仁. 车尔臣河年径流量气候成因及径流机制趋势分析[J]. 河南水利与南水北调, 2024(06): 27-28.