

20-26

滇西下关温泉水化观测项目的主成分分析 及有关地震异常研究

p315-7

唐采 唐吉阳
(云南省地震局, 昆明 650041)

摘要 根据对下关温泉1992年8个水化观测项目主成分进行的分析及影响因素的研究,发现氯和气压、水温和PH值、氩和氮、二氧化碳分别构成独立性较强的4个主成分,气压和流量是化学组分的主要影响因素。

关键词: 水化学测量 干扰识别 温泉 滇西 主成分分析

地震异常
地震预报

1 前言

滇西下关温泉位于下关西南约4 km处的东西向西洱河断层上,该点的某些观测项目(如水氩)已长达20余年,1988年以来,开展了氩、氮、氩、二氧化碳、水温、流量、酸碱度等多项水化观测项目,本文利用多元分析等方法对这些水化项目之间的关系进行了一些研究,目的是为消除干扰以便更有效地提取地震异常创造有利条件。

2 观测项目的主成分分析

为了分析多种观测项目之间的关系以及主要影响因素,我们使用多元统计方法之一的因子分析法,它主要是在关系复杂的情况下找到影响所研究的许多变量的共同因素和特殊因素,当许多变量之间存在一定相关关系时,设法找出若干个数目较少的新变量(它实质上是一些原始变量的线性组合),用这些独立的新变量(因子)来表达原始变量,找出影响观测项目的主要因素,这里我们采用的是研究变量之间关系的R型分析方法^[1]。

设样本数为 N ,每个样本有 v 个观测成分(变量), v 个成分可用 $x = (x_1, x_2, \dots, x_v)$ 这样一个向量来表示,如果我们分析了 $N \times v$ 个数据,则可用矩阵表示如下:

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1v} \\ x_{21} & x_{22} & \cdots & x_{2v} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nv} \end{bmatrix}$$

由 v 个单项变量线性组合构成少数几个综合指标(主成分),对综合指标的要求,一是要尽可能明显地反映出样本之间的差异,二是综合指标彼此之间线性无关,这样的综合指标能辨别原始变量中独立的不相关的线性组合,因此,需找一个 x 的线性函数,使得它在 x 的一切线性函

收稿日期:1996-04-30

第一作者简介:唐采,女,1954年5月出生,工程师,从事地震分析预报及监测工作。

数中有最大的方差.为此,通过正交变换使相关关系较高的变量(观测项目)变成新的一组变量 Y_j (原始变量的线性组合)

$$Y_j = \sum_{k=1}^v W_{jk} x_k \quad j = 1, 2, \dots, m$$

Y_j 是所求的主成分, W_{jk} 是第 j 个主成分对第 k 个变量的权系数, x_k 是第 k 个原始变量.变换前后总方差保持不变,但可使 Y_1 在总方差中所占比例最大(应用方差极大准则), $Y_2, Y_3, \dots, Y_m$ 的方差依次递减.一般按累积百分比 $\geq 85\%$ 挑选前 m 个主成分.具体计算步骤参见文献[1].

我们挑选了下关温泉观测项目中的流量(Q)、水氧(R_n)、水温(T)、二氧化碳(CO_2)、氩(Ar)、氮(N_2)、酸碱度(PH)和气压(P)作为变量,并利用1992年各项目的72个五日均值的数据(样本)进行了计算.首先求出各变量的算术平均值和标准差,再根据相关系数公式求出变量之间的相关矩阵 $R = (r_{jk})$

$$r_{jk} = \frac{\sum_{i=1}^N (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^N (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^N (x_{ik} - \bar{x}_k)^2}}$$

式中 i 为样本编号; $j, k = 1, 2, \dots, v$ (变量数).所求相关系数矩阵见表1.由于是对称矩阵,故只列出上三角部分.从表1中粗略可知各变量之间的关系.

表1 下关温泉1992年观测项目相关系数矩阵

$N=72$ $r_{0.05} = 0.232$

	流量	水氧	水温	二氧化碳	氩	PH值	氮	气压
	1	2	3	4	5	6	7	8
1	1.0000	-0.5145	0.1927	-0.6701	0.1499	0.3894	0.4133	-0.7454
2		1.0000	-0.1014	0.2990	0.1720	-0.1523	-0.2565	0.5251
3			1.0000	-0.0814	0.2272	0.6167	0.2236	-0.0570
4				1.0000	-0.0792	-0.2223	-0.2124	0.4249
5					1.0000	0.3139	0.5919	-0.1453
6						1.0000	0.3946	-0.1625
7							1.0000	-0.6099
8								1.0000

在相关系数矩阵(R)的基础上,用雅可比方法求其特征值(λ)以及与之相对应的特征向量 U .将 λ 值由大到小排列,按累积百分比 $\geq 85\%$ 的原则,取前4个作为主成分列于表2.在表2的基础上计算出各变量在各个主成分中的权系数(负荷),即求初始因子载荷矩阵 $A = (a_{ij}) = (\sqrt{\lambda_j} U_{ij})$,这里 $j=1, 2, 3, 4$, 属选取的因子个数.然后用方差最大准则旋转矩阵 A 后,计算出旋转后的因子系数矩阵(表3).根据表3作出图1.可直观地看出各观测项目之间有下列关系:

(1) 由正交旋转后的因子负荷矩阵可知,主成分①中各变量因子负荷,从正方向看,主要是氮和气压,其负荷值分别为0.8861和0.7453,它们的关系密切.水氧与气压的关系也遵循气体特性,因为氮也是一种气体,其溶解度受温度、压力的影响^[3].主成分分析结果证明了它

们之间确实存在密切关系.在负方向上,流量(Q)和氮(N_2)与主成分①亦有密切关系,唯负荷值较小而已(见图1(a)).

表2 主成分的特征值(λ)及累积百分比

主成分	λ	累积值	累计百分比	变量组合
①	3.295	3.295	41.18	主要为 Rn 和 P (气压)
②	1.667	4.962	62.03	主要为 PH 和 T
③	1.127	6.089	76.11	主要为 Ar 和 N_2
④	0.790	6.879	85.99	主要为 CO_2 (气)

表3 正交旋转后的因子负荷矩阵

观测项目*	主成分① $\lambda = 3.295$	主成分② $\lambda = 1.667$	主成分③ $\lambda = 1.127$	主成分④ $\lambda = 0.790$
1	-0.5345	-0.1899	0.1867	-0.7162
2	0.8861	-0.1180	0.1469	0.1299
3	-0.0430	0.9116	0.0712	0.0114
4	0.1284	-0.0657	-0.0260	0.9466
5	0.2091	0.1810	0.8875	-0.0706
6	-0.0810	0.8428	0.2331	-0.1953
7	-0.4163	0.1691	0.8230	-0.0796
8	0.7453	0.0580	-0.3812	0.3850

* 观测项目编码同表一

(2) 主成分②中因子负荷值最大的是水温(T)和 PH 值,分别为 0.9116 和 0.8428,在它的负方向上无重要的项目与之有关.它代表了该温泉中一组独立的变量,与其它的项目之间显示不出满足统计意义的密切关系(见图1(b)).

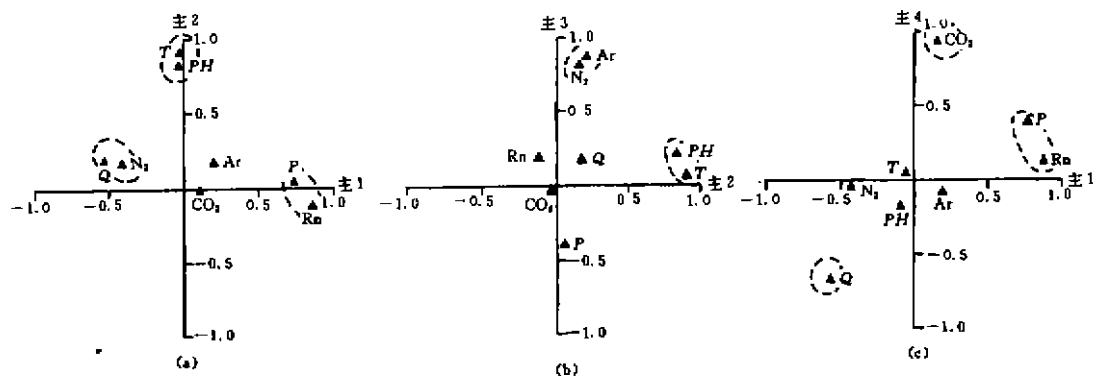


图1 下关温泉水化观测项目关系图

Fig.1 Relation map of Hydrochemical observations in Xiaguan hot spring.

(a) 以主成分①与②为坐标轴; (b) 以主成分②与③为坐标轴; (c) 以主成分①与④为坐标轴

(3) 主成分③中的主要变量是氢和氮,负荷值分别为 0.8875 和 0.8230,负方向上与之相

关的项目只有气压(P)。氡和氮的含量在有些年份其相关系数甚高,它们同源的可能性很大。

(4) 主成分④中主要起作用的是二氧化碳(CO_2), 负荷值为 0.9466. 负方向上的主要相关项目是流量(Q). 该主成分虽然权系数($\lambda = 0.790$)较小, 但仍反映了一组相对比较独立的成分. 从其与温度、 PH 值及其它化学成分不相关的情况看, 它来自地表及空气的可能性小, 极可能是从地壳深部迁出^[2], 迁出时浓度均匀, 中途与流量不均匀的循环水体混合, 故水量大小构成气体二氧化碳浓度变化的直接原因。

上述现象表明, 该温泉 8 个观测项目中的 3 个参量, 即流量、水温、 PH 值, 以流量的作用最为突出, 它虽被排除在 4 个主成分之外, 但它与氡及二氧化碳的负向关系甚为突出, 是研究水氡及二氧化碳动态变化时不可忽视的因素. 水温和 PH 值是相对较独立的一组变量, 与其它观测项目的密切程度不甚突出. 4 种化学成分中氮、氡构成密切程度较高的一组独立组合, 它可能代表了主要来自大气并经浅层水循环而带入的一组组分^[2]. 二氧化碳代表了该处地壳深部的组分. 而氡, 从其衰变规律看, 不可能来自深部, 它的产生深度可能介于二氧化碳(气体)与氮、氡所处地质环境之间。

值得注意的外界因素是气压, 它不但与水氡密切相关, 且与流量和氮有着负向的密切关系. 气温与气压的关系人所共知, 为了减少问题的复杂性, 未将其选入分析的参量之中。

3 相关因素的滞后影响研究

观测项目之间的相关关系, 已从前述资料中可见一斑, 但那是用同相位点计算的结果. 若将某一项目作为前导项目, 然后计算其它项目滞后数据与之相关的相关系数, 可大致了解有关项目滞后变化的特点。

我们主要研究水中化学元素的动态变化及其影响因素, 故将其物理参量(流量、水温)及外界因素气压作为影响因素研究. 氡气属放射性物质, 由于它是我们日常观测的重点, 这里既暂将其列入化学元素中, 同时也将其作为影响因素来研究. PH 值是水中氢离子浓度的反映, 此处暂将其归入化学元素. 这样, 研究的主要目标应是氡、氮、二氧化碳、氡和 PH 值, 而影响因素是流量、水温、气压和氡. 将影响因素作为前导项目, 分别计算滞后(1)点、(2)点和(3)点的相关系数, 将这 3 点中最大的相关系数列出. 表 4 是利用 1992 年五日均值资料计算的结果, 其中带 * 号的部分表示: ①满足 95% 的信度; ②与表一比较, 相关系数有所改善和提高(绝对值大于 0.1)。

表 4 化学元素滞后相关系数

$N=60$, $r_{0.05}=0.232$

前导项	滞后项				
	Rn	CO_2	Ar	PH	N_2
Q	0.31 (1)	-0.67 (1)	0.40* (1)	0.48* (1)	0.58* (1)
Rn		0.16 (1)	-0.28* (1)	-0.44* (1)	-0.36* (1)
T	-0.12 (3)	-0.20 (1)	-0.04 (2)	-0.04 (2)	0.15 (1)
P	0.75* (1)	0.52* (3)	-0.08 (3)	-0.14 (3)	-0.44 (1)

注: (1)、(2)、(3)代表滞后点数

由于使用的是五日均值, 滞后 1 点约相当于 5d 时间. 故从表 4 可以看出, 流量仍是本点的

主要影响因素,它虽对水氧、二氧化碳滞后的影响不大,但对氩、氮及 PH 值滞后 5d 的影响最大,且均为正相关;氧对氩、氮及 PH 值的滞后影响与流量类似,但为负相关;水温对化学元素的滞后影响不显著;气压对水氧和流量滞后 5d 的影响最大,相关系数分别为 -0.79 和 0.75 ,对二氧化碳滞后 15d 的影响相当明显.而流量和水氧对氩、氮及 PH 值、水温滞后影响同步,其中可能包含了气压的影响.水温和 PH 值互相关程度较高(表 1),它们对其它项目影响不显著,似乎只是个被影响角色,且具一定的独立性.氩、氮的互相关程度亦较高,它们与水温及 PH 值不相关,但对二氧化碳有滞后 5d 的影响,相关系数分别为 -0.44 和 -0.62 .

综合表 1 和表 4 及有关计算结果,将观测项目之间的相关关系及滞后影响构成系统(图 2).图中给出了 8 个观测项目明显表现出的 4 个层次.气压是第一层次,似乎是干扰源.流量和水氧构成第二层次,它们之间同步变化,但同时受第一层次的干扰,且又影响下一层次.第三层次存在互相独立的两组,即水温和 PH 值,氩和氮,两组之间互不相关,各组内互相关程度很高.第四层次以二氧化碳为代表,它对别的项目影响不显著,但却受着气压、流量、氩、氮等多种组分的影响.对比因子分析结果和滞后相关分析结果均与其基本相同.

4 多变量复合异常初步研究

研究地震异常最常用的方法是先对单项前兆进行有关处理(如排除干扰等),然后再对异常及其相应地震进行研究.但从前述分析看,有的项目影响因素很多,且滞后影响的时间长短不一,进行数据处理相当烦琐.有的项目影响因素虽少(如流量),但又不是重点观测的成分,且有时与其它成分存在互相关关系,属主属次难以区分.前述主成分分析结果表明,下关温泉 8 种观测项目可区分成 4 种独立的主成分,这为用主成分研究异常提供了方便.

主成分是一种观测项目的组合,特点是主成分内项目之间相关程度较高,但与其它主成分的项目相关程度较弱或不相关.主成分内的项目成因机制类似或有一定的同源性(如氩和氮),不同主成分的项目可能源于不同的机制.假如地震前各种主成分同时经受地壳应力的作用,主成分内的项目同步变化并不稀奇,而不同主成分的项目应该表现出一定程度的同步变化.根据这一想法,我们采用加和比方法探索异常.计算之前,应对原始数据进行统一量纲的处理.这里首先将数据标准化:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{d_j}, \quad \bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$$

$$d_j = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{ij} - \bar{x}_j)^2}$$

加和比公式如下:

$$D = \sum_{j=1}^{P_1} x'_{ji} / \sum_{j=1}^{P_2} x'_{ji} \quad i = 1, 2, 3, \dots, N; \quad j = 1, 2, 3, \dots, P_1, P_2$$

N 为样本总数, P_1 和 P_2 分别为分子和分母的变量数.计算结果 D 是复合项目的比值,我们选择了 4 种组合的加和比值(D)曲线(图 3).其中曲线(a)为 $(Q + CO_2)/(Rn + Ar)$ 的比值, (b)为

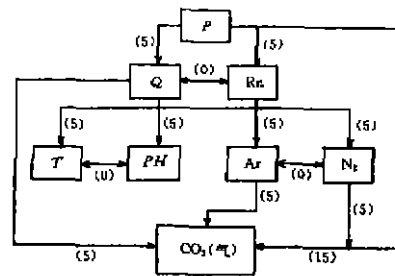


图 2 下关温泉观测项目相关及滞后相关系统

Fig. 2 Correlation and lag correlation of observation items in Xiaguan hot spring.

→ 滞后影响; ↔ 互相关; () 滞后天数

($T + CO_2 + PH$)/($Ar + N_2 + P$)的比值,(c)为($Q + Rn + T + CO_2$)/($Ar + PH + N_2 + P$)的比值;(d)为($Q + Ar$)/($Rn + CO_2$)的比值。

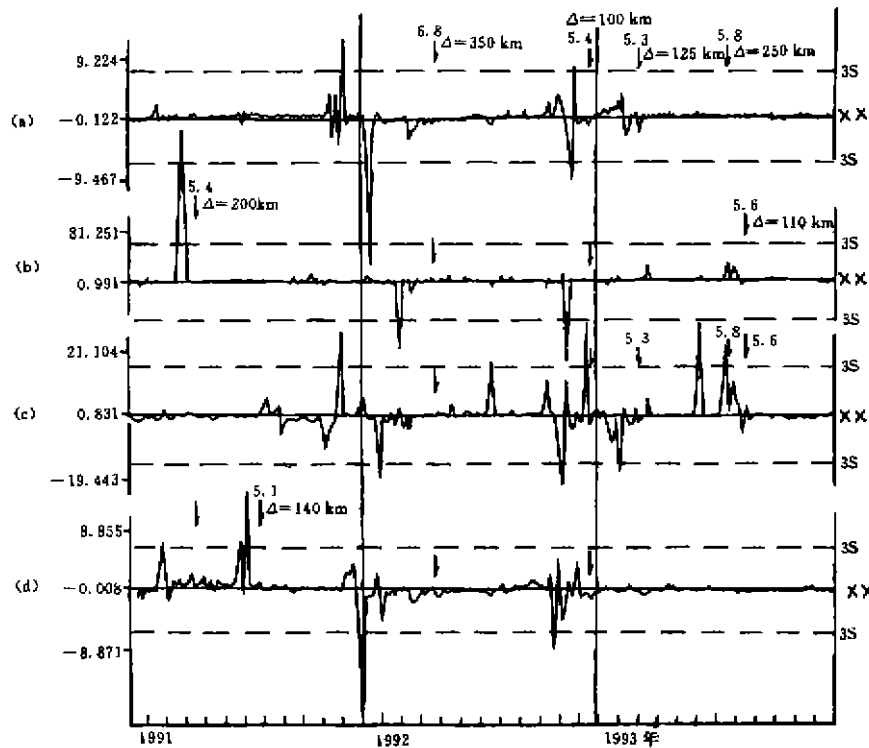


图3 下关温泉水化观测项目加和比值曲线(说明见本文)

Fig.3 Additivity ratio values curves of hydrochemical observation items in Xiaguan hot spring.

总体看来基值比较稳定,异常突出,但各曲线之间仍有区别,如曲线(c)由于选择的项目较多,地震前表现多为多点突跳,其它曲线选择的项目较少,异常多为单点突跳.若以数据的3倍标准差为异常界线,有两个时间段4条曲线全部出现异常,一个是1991年11月到1992年2月,4个月内相继出现多次突跳异常,反映一次地壳应力加强过程,但距该点300 km以内无大于6级地震.1992年4月23日澜沧西中缅边界连续发生6.8和6.7级地震,距该点350 km,此段异常有可能对应了这两次地震;另一个异常段是1992年10月中旬到12月上旬,4条曲线均出现突跳异常,曲线(d)异常最早,曲线(a)最晚,曲线(c)异常幅度最大.该异常是1992年12月18日永胜5.4级地震的反映,震中距仅100 km.初步统计,4条曲线出现非同步的突跳异常共7次,有6次在异常结束后一个半月以内(距该点200 km以近)对应了5.1~5.6级地震,对应率在80%以上,比单项目对应率高出很多.

从单条曲线比较看,曲线(c)选了所有的观测项目,突跳现象复杂,有虚有漏,不甚理想.曲线(a)选择项目较少,无虚报但漏报较多,亦不理想.曲线(b)异常亦是无虚有漏,它对宁蒗5.4级地震反映明显,对姚安地震略有反映,对应效果比曲线(a)略好.曲线(d)异常无一虚报,对应地震也较多,相对较好.

5 结束语

经对下关温泉水化观测项目的统计分析,可得如下几点认识:

(1) 8个观测项目可组合成4个主成分,即①气压和水氨、②水温和 PH 值、③氩和氮、④气体二氧化碳.主成分内各项之间是强相关,主成分之间相关程度很弱或基本不相关.前两个主成分的 λ 值(特征值)占整个 λ 值的 62%,似乎应受到更多的重视.

(2) 主要的影响因素是气压和流量,它们对其余项目都有着滞后影响.水温和 PH 值主要受主成分①的影响,它们对其它项目影响不明显.气体二氧化碳受多种因素的影响,由此推断它可能包含有由循环水带入的地表成分,是否全是深部成因,尚需进一步证实.

(3) 用不同主成分的观测项目搭配,进行相加后再相比,构成复合项目,其异常在基值稳定的基础上,显明直观,对应效果较好.若能综合考虑多条曲线的异常,这将对距该点 200 km 以内的 5 级以上地震起到很好的监视作用.

由于所用资料有限,本方法对强震的监视作用有待进一步研究.

参考文献

- 1 中国科学院地质研究所.数学地质引论.北京:地质出版社,1978.178.
- 2 国家地震局地质研究所,云南省地震局.滇西地区活断裂.北京:地震出版社,1990.162.
- 3 国家地震局分析预报中心三室.地下流体预报地震论文集.北京:地震出版社,1987.81.
- 4 中国科学院计算中心概率统计组.概率统计计算.北京:科学出版社,1979.

ANALYSES ON MAIN COMPOSITIONS OF HYDROCHEMICAL SURVEY ITEMS IN XIAGUAN HOT SPRING OF WEST YUNNAN AND STUDIES ON ANOMALIES RELATED TO EARTHQUAKES

TANG Cai TANG Jiyang

(*Seismological Bureau of Yunnan Province, Kunming 650041*)

Abstract

In this paper, main compositions of 8 hydrochemical observation items in Xiaguan hot spring in 1992 have been analysed and the disturbance factors have also been studied. The results show that radon and air pressure, water temperature and PH values, argon and nitrogen, and CO_2 are 4 main compositions which have stronger independence. Air pressure and flow quantity are the main disturbance factors to the chemical compositions.

Key words: Hydrochemical survey, Interference recognition, Hot spring, West Yunnan, Main composition analysis