

降水随机模拟中的超离差现象

R. W. Katz

(美国国家大气研究中心)

M. B. Parlange

(美国约翰·霍普金斯大学)

摘要 简单的随机模式拟合的日降水量时间序列具有过低估计月(或季)总降水观测(或年际)变化的显著趋势。研究发展一种称为链相关过程的特殊类型的随机模式,估计了“超离差(overdispersion)”现象的程度。这种现象可归因于模式对降水的高频变化考虑不充分。对加利福尼亚奇科(Chico, California)1月份日降水量,拟合更复杂的随机模式,大大地减少月总降水方差的过低估计。通过利用日降水出现的高阶马尔科夫(Markov)链能够完全消除超离差的一种来源(降水日数)。然而观测记录的某些方差仍不能得到解释,它可能归因于低频变化,有时称“潜在可预报性(potential predictability)”。特别有意义的事实是这些更复杂的随机模式仍然低估月方差,比可选择方法(受大尺度大气环流指数条件限定的最简单的链相关过程形式)低估的还多。

关键词 降水 随机模拟 超离差现象

引言

用简单的随机模式拟合日降水量时间序列时,有低估月(或季)总降水观测记录(或年际)方差的显著趋势,已为大家所熟知(Buis-hand, 1978; Wilks, 1989)。在统计文献中,观测记录方差超过拟合模式方差的这种情况,称为“超离差”(Cox, 1983)。但对降水超离差现象的解释是不一致的。有些研究者将这种偏差看作是模式对降水的高频变化考虑不充

分的证据(Gregory 等, 1993)。另一些研究者将它归因于低频变化,而这些模式并没有进行说明,照此应制定月总降水“潜在可预报性”年际时间尺度的度量(Shea 和 Madden, 1990; Shea 等 1995; Singh 和 Kripalani, 1986)。但是,若第一种解释是有根据的,那么,这种方法将过高估计潜在可预报性的程度。

在本文中,我们考查第一种解释,利用称为链相关过程(chain-dependent process)

(Katz, 1977a)的降水高频变化的一种特殊类型的随机模式,识别和消除低估随机模式方差的多种来源。这种模式包含将降水过程分为两个分量模式,一是降水出现,一是降水强度(即有关降水出现的降水量),将考虑这些模式的有限外延,其中包括日降水出现过程的高阶马尔科夫链和动态降水强度的自相关。在降水随机模拟文献中,有支持进行这种调整的论据,但其重点不在于月总降水方差近似如何好。因而,本文的方法将填补降水日方差与月方差之间关系的研究。

第一部分研究日降水量时间序列随机模式的一些性质。对本项研究要点是称为“随机总和”的月总降水量的表达式,能使其方差分解为两项,其中一项包含降水日数的方差。第二部分将这些理论结果应用于加利福尼亚州的奇科 1 月份日降水量时间序列。Katz 与 Parlange (1993, 1996) 以前分析过同一资料集,进行了大尺度大气环流对局地降水的统计减尺度研究。这些研究提高了奇科更复杂的高频变化降水模式减少超离差现象扩展的可能性。通过综合分析表达式、计算方法和模拟技术,对链相关过程的多种外延,估算了 1 月份总降水方差与有关的统计。一些技术细节列在附录中说明(略)。最后,第三部分是讨论。

1 随机模式

本节利用总降水的概率表达式以随机总和的形式表示总降水。不严格地说,基本概念是,月总降水是由单个风暴所形成的降水量的总和组成。这种表示法使总降水的方差与日降水量特定的随机模式的多种分量相关。由于这个原因,这种方法比仅是降水链相关过程的最简单形式具有更大通用性的优点(Katz, 1977a)。

1.1 随机总和的表示

长度为 T 日的某时段(例如一个月)的总降水,记为 $S(T)$,可表示为

$$S(T) = Z_1 + Z_2 + \cdots + Z_{N(T)} \quad (1)$$

式中,在该时段内降水出现的次数(即降水日数),以 $N(T)$ 表示,它本身是随机变量。 $Z_k > 0$, 表示第 k 次降水出现的强度(即第 k 个降水日的降水量), $k = 1, 2, \dots, N(T)$, 定为独立的恒等分布,并具有平均值 $\mu = E(Z_k)$ 与方差 $\sigma^2 = \text{Var}(Z_k)$ 。 $N(T)$ 的计算过程中假设 Z_k 独立,以 π 表示给定日降水出现的非条件概率(即“计数”的出现率)。为了简便以及因为仅关心持续一个月的时段,这些参数的任何年周期均被忽略。

由于以降水日数为条件,月总降水方差表示为

$$\begin{aligned} \text{Var}[S(T)] &= E\{\text{Var}[S(T) | N(T)]\} + \\ &\quad \text{Var}\{E[S(T) | N(T)]\} \\ &= E[N(T)]\text{Var}(Z_k) + \\ &\quad \text{Var}[N(T)][E(Z_k)]^2 \\ &= T\pi\sigma^2 + \text{Var}[N(T)]\mu^2 \end{aligned} \quad (2)$$

第一种表达方式实际包含任意两个随机变量(Lindgren, 1968, 118), 第二种表达方式特指随机总和(Feller, 1968, 第 12 章), 第三种表达方式利用某些简化符号。因而,(2)式把总降水方差的组成分解为两项,一项相应于强度 T_x 的普通(即非随机)总和的方差(注意:降水日的期望数是 $E[N(T)] = T\pi$), 另一项包含降水日数的方差 $\text{Var}[N(T)]$ 。

1.2 链相关过程

链相关过程(Katz, 1977a; Todorovic 和 Woolhiser, 1975)具有合乎需要的特点,即只需若干相对少的参数,就能表明降水时间序列的一些最重要的统计特征。其简单的结构能分析确定包括月总降水方差(Katz, 1977b)在内的许多性质。在 1.1 节特别引入了应用于这类随机模式总降水随机和的表达式。

雨期(即在该期降水出现的动态雨日过程)或干期持续的趋势采用日降水出现的一阶马尔科夫链模式两种状态表示。令 $\{J_t; t = 1, 2, \dots\}$ 表示日降水出现序列(即 $J_t = 1$ 表示 1 雨日, $J_t = 0$ 为 1 干日)。这种模式的特点是具有转移概率

$$P_{ij} = P_r\{J_{t+1} = j | J_t = i\} \quad i, j = 0, 1 \quad (3)$$

具有约束: $P_{i0} + P_{i1} = 1, i = 0, 1$ 。

依据雨日的概率(如 1.1 节所述,采用 $\pi = P_r\{J_t = 1\}$ 表示)以及一价自相关系数(或持续参数), $d = \text{Corr}(J_t, J_{t+1})$, 对再参数化马尔科夫链是方便的。 π 与 d 这两个参数和转移概率有关

$$\begin{aligned} \pi &= P_{01}/[1 - (P_{01} - P_{11})] \\ d &= P_{11} - P_{01} \end{aligned} \quad (4)$$

注意对于日降水出现的时间序列, $d > 0$ 。

雨日数能表示为总和 $N(T) = J_1 + J_2 + \dots + J_T$ 。它的方差,如同出现在总降水方差的总表达式(2),在一阶马尔科夫链(3)式情况下,能近似为(对大日数 T):

$$\begin{aligned} \text{Var}[N(T)] &\approx \\ T\pi(1 - \pi)[(1 + d)/(1 - d)] \end{aligned} \quad (5)$$

显然,由于(5)式右边出现因子 $(1 + d)/(1 - d)$, 马尔科夫相关使与独立性有关的方差(即二项式分布的方差是 $T\pi(1 - \pi)$ 的方差膨胀(若 $d > 0$)。雨日数的精确方差能通过 $N(T)$ 准确分布的递推计算确定(Katz, 1974)。

如 1.1 节连同一般随机总和所已经定义的那样,假设日降水强度 Z_k 链相关过程为独立恒等分布。引进另一符号是方便的,令 $\{X_t: t=1, 2, \dots\}$ 表示日降水量时间序列(即 X_t 假设为 0 与正值)。强度 $X_t > 0$ (即在 $J_t = 1$ 的 t 日)的相当假设是条件独立恒等分布,给定日降水出现的马尔科夫链模式的状态。特别就 X_t 而论,强度平均数与方差能定义为 $\mu = E(X_t | J_t = 1)$ 与 $\sigma^2 = \text{Var}(X_t | J_t = 1)$ 。日强度具有正偏斜分布, Todorovic 和 Woolhiser (1975) 定为指数分布, Katz 定为伽马分布(1977a), Katz 和 Parlange (1993) 根据乘方变换为正态分布。

X_t 过程的非条件平均数、方差和自相关函数与强度平均数和方差以及马尔科夫链的参数有关:

$$\begin{aligned} E(X_t) &= \pi\mu \\ \text{Var}(X_t) &= \pi\sigma^2 + \pi(1 - \pi)\mu^2 \\ \text{Corr}(X_t, X_{t+1}) &= \pi(1 - \pi)\mu^2 d / \text{Var}(X_t) \end{aligned}$$

$$l = 1, 2, \dots \quad (6)$$

显然, X_t 过程的自相关是通过马尔科夫链的自相关导出(即 $\text{Corr}(J_t, J_{t+1}) = d^l, l = 1, 2, \dots$, 出现于(6)式)。

当然,月总降水也有普通非随机总和的更方便表示, $S(T) = X_1 + X_2 + \dots + X_T$ 。那么,它的方差包含 X_t 过程的自相关总和(见附录 A 的(A1)与(A2),略)。或者将雨日数方差的近似表达式(5)代入(2)式或将自相关函数的表达式(6)代入(A2),就可得出大 T 总降水方差(Katz 和 Parlange, 1993)。

$$\begin{aligned} \text{Var}[S(T)] &\approx T(\pi\sigma^2 + \\ \pi(1 - \pi)[(1 + d)/(1 - d)]\mu^2) \end{aligned} \quad (7)$$

通过比较(7)式与相应的观测记录方差,就能获得超离差程度的估计。由于链相关过程拟合的日降水量将有效地重建日方差,所以只通过外延影响自相关函数,就能够增加估算的月总降水方差。

1.3 外延

1) 高阶马尔科夫链

确定日降水出现的随机模式是一阶以上的马尔科夫链是十分自然的(Chin, 1977; Gates 和 Tong, 1976)。例如,二阶马尔科夫链的特点是具有取决于前两日降水是否出现的转移概率:

$$\begin{aligned} P_{ijk} &= \text{Pr}\{J_{t+1} = k | J_t = j, J_{t-1} = i\} \\ i, j, k &= 0, 1 \end{aligned} \quad (8)$$

注意,一阶链是(8)式在 $P_{ijk} = P_{jk}, i, j, k = 0, 1$ 约束下的特殊情况。雨日 π 和持续参数 d (仍作为一阶自相关系数解释)的概率能从(8)式二阶转移概率导出(见附录 B, 略)。

由于具有向量状态一阶链的高阶马尔科夫链的空间状态表达式,使得这种模式与具有两种以上状态的一阶马尔科夫链是等价的。因此,如果 1.3 节所述的日降水量随机模式被拓展为一体化的高阶马尔科夫链,就能使用链相关过程更普遍的理论(Katz, 1977b)。特别是链相关过程总和方差的(7)式是包含矩阵转置公式的特殊情况,也是为 Klugman 和 Klugman(1981)所采用的方法。

再一次运用(2)式分解月总降水的方差也能开拓表示随机总和的(1)式。雨日数方差 $\text{Var}[N(T)]$ 的近似表达式(5)建立在一阶马尔科夫链基础上,需要改进。虽然(5)式不存在简单的模拟量,但能通过递推方法计算精确的方差。附录C(略)给出确定(8)式二阶马尔科夫链雨日数精确分布的算法,它是Katz(1974)给出方法的推广。期望高阶马尔科夫链将增加估算的雨日数方差,并因此增加估算的月总降水方差。

2) 自相关强度

另一方面,链相关过程的外延包含允许一定雨期内的强度自相关是缺乏普遍性的。形式上采用强度遵循自相关系数 $\varphi > 0$ 的一阶自回归[AR(1)]过程,当一雨期达终点时,这种过程随机结束。在Katz和Parlange的文献中(1995),AR(1)过程实际拟合为乘方转换的强度,同样允许偏斜度。他们利用这种方法模拟小时降水量时,其中强度自相关的这种情况更明显。先前也试图让强度间的相关依靠多状态马尔科夫链,这种方法是需要估算大量转移概率的(Gregory等,1993; Haan等,1976)。

如果出现过程的一阶马尔科夫链模式持有1.2节的形式,那么, X_t 过程的自相关函数(Katz和Parlange,1995)为

$$\text{Corr}(X_t, X_{t+1}) = [\pi\sigma^2(P_{11}\Phi)^l + \pi(1-\pi)\mu^2 d^l] / \text{Var}(X_t) \quad l = 1, 2, \dots \quad (9)$$

注意,根据(4)式(9)式的转移概率可表示为 $P_{11} = \pi + (1-\pi)d$ 。(9)式 $\text{Var}(X_t)$ 仍如(6)式所给出的。但现在(9)式的自相关函数是两项的加权平均,一项与强度自相关函数(即 Φ^l)有关,另一项与出现的自相关函数(即 d^l)有关。

对于大 T ,根据(9)式与附录(A2),月总降水方差近似为

$$\text{Var}[S(T)] \approx T\{\pi[(1+P_{11}\Phi)/(1-P_{11}\Phi)]\sigma^2 + \pi(1-\pi)[(1+d)/(1-d)]\mu^2\} \quad (10)$$

我们注意到,由于(10)式右边包含因子 $(1+P_{11}\Phi)/(1-P_{11}\Phi)$,增加了方差(若 $\Phi > 0$), (10)式概括了(7)式。由于出现于(1)式的强度假定是独立的,即使(10)式不再严格地包含随机总和,但它仍形成有点类似(7)式的方差分解。第一项相当AR(1)过程的非随机总和的方差,具有自相关系数 $P_{11}\Phi$,出现 P_{11} 因子是因为它支配雨期的随机结束。

3) 非恒等分布

日降水链相关过程的原公式实际考虑到另外的复杂情况,允许日强度分布取决于降水是否在前日出现(Katz,1977a)。确切地说,定义条件平均数和方差为

$$\begin{aligned} \mu_i &= E(X_t | J_{t-1} = i, J_t = 1) \\ \sigma_i^2 &= \text{Var}(X_t | J_{t-1} = i, J_t = 1) \\ & \quad i = 0, 1 \end{aligned} \quad (11)$$

除Katz(1977a)外,Chin和Miller(1980)以及Klugman和Klugman(1981)都用这种模式拟合日降水资料(均见Guttorp文献的第二章,1995)。

对于链相关过程的这种形式,已给出(Katz,1977a,b)日降水量的平均数、方差与自相关函数:

$$\begin{aligned} E(X_t) &= (1-\pi)P_{01}\mu_0 + \pi P_{11}\mu_1 \\ \text{Var}(X_t) &= (1-\pi)P_{01}(\sigma_0^2 + \mu_0^2) + \pi P_{11}(\sigma_1^2 + \mu_1^2) - [E(X_t)]^2 \\ \text{Corr}(X_t, X_{t+1}) &= [\pi(1-\pi)\mu(P_{11}\mu_1 - P_{01}\mu_0)d^{l-1}] / \text{Var}(X_t) \\ & \quad l = 1, 2, \dots \end{aligned} \quad (12)$$

注意,非条件强度平均数 μ (出现于(12)式)和方差 σ^2 可能与条件强度平均数 μ_i 与方差 σ_i^2 有关(见附录B(B3),略)。当 $\mu_0 = \mu_1$ 与 $\sigma_0^2 = \sigma_1^2$ 时,(12)式简化为(6)式。最后,对大 T ,月总降水量的近似方差可根据(12)式与(A2)的自相关函数导出

$$\text{Var}[S(T)] \approx T\langle \text{Var}(X_t) + [2\pi(1-\pi)\mu(P_{11}\mu_1 - P_{01}\mu_0)] / (1-d) \rangle \quad (13)$$

就两种条件的强度平均数与方差的差异程度

表 1 转移概率
(各阶马尔科夫链拟合加利福尼亚奇科 78 年 1 月份日降水出现的时间序列)

前几日状态				$Pr\{J_{t+1}=1 J_t, J_{t-1}, J_{t-2}, J_{t-3}\}$			
J_{t-3}	J_{t-2}	J_{t-1}	J_t	1 阶	2 阶	3 阶	4 阶
0	0	0	0	0.2109	0.1838	0.1691	0.1584
0	0	0	1	0.5705	0.5882	0.5767	0.5541
0	0	1	0	0.2109	0.3105	0.2541	0.2857
0	0	1	1	0.5705	0.5576	0.5806	0.6083
0	1	0	0	0.2109	0.1838	0.2488	0.2184
0	1	0	1	0.5705	0.5882	0.6344	0.5161
0	1	1	0	0.2109	0.3105	0.3523	0.3188
0	1	1	1	0.5705	0.5576	0.5415	0.4857
1	0	0	0	0.2109	0.1838	0.1691	0.2039
1	0	0	1	0.5705	0.5882	0.5767	0.5918
1	0	1	0	0.2109	0.3105	0.2541	0.2000
1	0	1	1	0.5705	0.5576	0.5806	0.5254
1	1	0	0	0.2109	0.1838	0.2488	0.2679
1	1	0	1	0.5705	0.5882	0.6344	0.7167
1	1	1	0	0.2109	0.3105	0.3523	0.3922
1	1	1	1	0.5705	0.5576	0.5415	0.5882

来说, (13) 式的方差将超过普通链相关过程 (7) 式的方差。

4) 外延 1), 2) 与 3) 的组合

链相关过程外延三种类型的多种组合自然能同时运用。如前所述, 链相关过程的一般理论 (Katz, 1977b) 已包含外延 1) 与 3) 两方面以及它们的组合 (即出现与非恒等分布强度的高阶马尔科夫链)。虽然月总降水量方差存在闭合形式的表达式, 但它包含大状态空间和矩阵转置 (Klugman 和 Klugman, 1981)。包括所有三种外延的理论也已提出, 在附录 A (略) 仍说明如何导出在一般情况下日降水量的一阶和二阶自相关系数。

2 结果

分析了加利福尼亚州奇科 78 年 (1907~1988 年, 因缺测除去 4 年) 1 月份日降水量的时间序列。由于奇科的记录时间长, 能获得 1 月份总降水量相当可靠的年际方差估算值。奇科位于美国西海岸附近, 冬季大尺度大气环流形势对这个地区局地天气的影响是主要的。因而降水过程应该相对持续, 基于此, 进

行了有说服力的简单随机模式试验。对于这些资料, Katz 与 Parlange (1993) 已经证实, 普通链相关过程具有一定程度的超离差。

2.1 拟合模式

分别提出日降水出现与强度的拟合模式。为此重申我们的重点不在是否需要更复杂的模式提供改进的拟合 (即通过认为“统计上有意义”的参数估算), 而在这些模式估算 1 月份总降水方差和相关统计的能力上。

1) 降水出现的处理

表 1 给出两种状态 1~4 阶马尔科夫链拟合奇科 1 月份日降水出现时间序列估算的转移概率 (基于近似最大相似性准则)。由于对转移概率约束 (如 1.2 节所述), 仅列出雨日的条件概率。

为了便于比较且由于任一低阶链都能看作是较高阶链的特殊情况 (如 1.3 节所解释), 表中列出了对应于四阶链各阶的估算。对表 1 的每一行, 如果降水出现的时间序列实际由一阶链形成, 转移概率将是常数 (抽样误差除外)。表中列出一些有关于期呈持续形式不能通过一阶链模拟的证据。例如, 估算一

表 2 各种模式拟合加利福尼亚奇科 1 月份日降水强度参数估计值(787 个雨日)

模式形式		平均数		标准差		自相关
恒等分布	自相关	μ_0/mm	μ_1/mm	σ_0/mm	σ_1/mm	Φ
是	否	13.36	13.36	14.68	14.68	0
是	是	13.36	13.36	14.68	14.68	0.161
是	膨胀	13.36	13.36	14.68	14.68	0.411
否	否	11.62	14.84	12.15	16.28	0
否	是	11.62	14.84	12.15	16.28	0.145
否	膨胀	11.62	14.84	12.15	16.28	0.298

雨日的条件概率,仅从前日是干日的约 0.211,降到持续 4 日是干日的约 0.158(这两种概率对一阶链应是相同的)。对湿期这种形式还不够明显。

对于一阶马尔科夫链来说,转移概率估算(通过(4)式)能变换成相应的雨日与持续参数 $\hat{\pi}=0.3293$, $\hat{d}=0.3596$ (此处“ $\hat{}$ ”表示参数估算值)的非条件概率估算。当阶增加时 π 和 d 在数值变化上,有少许差别(例如二阶链 $\hat{\pi}=0.3290$ 和 $\hat{d}=0.3603$,根据附录 B(略)的方法确定)。这种情况有点类似自回归过程。在自回归过程中,高阶模式,仅通过较大滞后的自相关调整实现改进拟合。

2) 降水强度处理

表 2 给出奇科 1 月份日降水强度的条件平均数,条件标准差与一阶自相关系数的估算参数。为了方便,在一般形式上列出非恒等强度分布的这些参数估计值,回忆恒等分布情况相应的约束 $\mu_0 = \mu_1$ 与 $\sigma_0^2 = \sigma_1^2$ (见 1.3 节)。显然根据表 2,假定前一日是与雨日相反的干日,估计的条件强度平均数与标准差较小。为了便于比较月总降水方差的估计值,采用动差方法估计这些参数。选择的方法是将参数估算建立在正态性乘方转换基础上,并考虑到日强度的正偏斜分布(Katz 和 Parlange, 1993)。但是,利用非线性变换导致某种程度偏离,意味着采用这种技术未必就能确切重建原未转换的日强度样本平均数和方差。考虑这种偏斜度只有在研究月总降水分布形状(不仅是其方差)时才有意义。

表 2 的日强度一阶自相关系数 Φ 的估

计值是相对小的正值 0.161 或 0.145,它取决于强度分布是否假设为恒等分布。更一般而言,估算的自相关对取决于雨期内状态(雨期的第 1 日与第 2 日的比较等)的强度分布对任何差异十分敏感,非恒等分布模式的改进在 1.3 节描述。为了解决这问题,基于重建日降水量样本的一阶自相关系数选择估算 Φ 的方法也包含在表 2 中。在恒等强度分布情况下,这种方法包含将 μ 与 σ 估计值(与一阶马尔科夫链 π, d 与 P_{11} 一起)以滞后 $l=1$ 代入(9)式右边,并令该式与日降水量样本一阶自相关系数 0.279 相等,然后求 Φ 。在非恒等分布情况下,采取同样方法,但是将 μ_i 和 σ_i^2 , $i=0,1$ (连同一阶马尔科夫链 π, P_{01} 和 $P_{011}=P_{111}=P_{11}$)代入(A4)。获得稍大的 Φ 的估计值,在表 2 中称‘膨胀’:对恒等分布而言,是 0.411 和 0.161;对非恒等分布而言,0.298 和 0.145,膨胀程度稍小。非恒等分布的更复杂模式(即明确考虑雨期内的状态)能全部消除这些误差。

2.2 超离差的估算

对日降水量时间序列利用这些随机模式,能估算月总降水量的方差。由于这些模式的诊断能力,也能分析雨日数的方差和日降水量的自相关函数。为了获得这些统计的估计值,使用公式计算方法或随机模拟(建立在 10 000 a* 形成的 1 月份日降水量基础上)。所有这些方法都需要拟合模式参数的数值。我们仅代入相应的参数估计值(在 2.1 节给

* 原文可能有误——编者。

表 3 加利福尼亚奇科 1 月份降水超离差估算与相关统计
(根据日随机模式,沿用相应的观测值参数估计见表 1、表 2)

模式形式			导出的统计值			
马尔科夫链阶	恒等强度分布?	强度自相关?	雨日数标准差	自相关量(滞后 1 日)	自相关量(滞后 2 日)	总降水量(标准差/mm)
1	是	否	3.76	0.128	0.046	68.7
1	是	是	3.76	0.187	0.052	71.8
1	是	膨胀	3.76	0.279	0.082	77.6
1	否	否	3.76	0.160	0.058	71.4
1	否	是	3.76	0.218	0.062	74.5
1	否	膨胀	3.76	0.279	0.078	77.3
2	是	否	4.00	0.129	0.065	71.1
2	是	是	4.00	0.188	0.071	74.5
2	是	膨胀	4.00	0.279	0.100	79.5
2	否	否	4.00	0.160	0.073	73.2
2	否	是	4.00	0.217	0.078	76.4
2	否	膨胀	4.00	0.278	0.093	78.4
3	是	否	4.23	0.129	0.065	73.5
3	是	是	4.23	0.189	0.071	76.9
3	是	膨胀	4.23	0.281	0.100	81.0
3	否	否	4.23	0.160	0.073	76.5
3	否	是	4.23	0.218	0.078	79.5
3	否	膨胀	4.23	0.279	0.093	82.5
4	是	否	4.42	0.129	0.068	75.4
4	是	是	4.42	0.189	0.073	79.4
4	是	膨胀	4.42	0.281	0.103	83.7
4	否	否	4.42	0.160	0.076	79.0
4	否	是	4.42	0.218	0.080	80.6
4	否	膨胀	4.42	0.278	0.095	83.9
观测记录			4.33	0.279	0.113	88.6

出),忽略这些估计值的样本误差。大数目的日降水观测记录(2418=78×31)表明这样的误差是相对小的。

1) 雨日数

鉴于(2)式方差的分解,我们现在的重点在于如何适当通过日降水出现的马尔科夫链模式找到奇科 1 月份雨日数观测记录的标准差。表 3 列出了 1~4 阶马尔科夫链(参数估计值见表 1)估算的雨日数标准差。这些估算的标准差通过附录 C(略)概述的计算方法获得。

对于一阶马尔科夫链,雨日数的标准差估计值为 3.76 d(由近似表达式(5)得出几乎相同的值,为 3.81)。远低于观测记录值 4.33 d,或方差约 25%的超离差(表 3)。随着链阶

增加,估算的标准差也增加,三阶链稍低于观测记录值,而四阶链则略高。四阶链的这种高估值很可能不反映实际的“低离差”(underdispersion),而是观测记录和模式估算标准差两者的样本误差。

最重要的是通过利用高阶链已完全消除雨日数的超离差现象。不足的是当阶数增加时,需要估算的转移概率数目迅速增加(例如四阶链为 16 个参数,见表 1)。这样,可能出现过度拟合。还有一种方法(这里不详谈)是拟合一个高阶链更精确的模式,其中参数间给予一定的约束((Raftery,1985)。

2) 自相关

在月总降水方差的(A1)与(A2)表示方面,我们现在的重点在于如何适当采用多种

形式的随机模式,重建日降水量自相关函数。表 3 列有利用(A4)与(A5)计算的滞后 $l=1$ 和 2 d 的自相关系数。链相关过程的普通形式具有低估日降水量一阶自相关系数的显著趋势(即表 3 中 0.128 对观测记录值 0.279)。允许非恒等分布强度或自相关强度会稍增加这一估计值,两者组合产生一个值 0.218,仍远低于观测记录值。必然地,‘膨胀’强度自相

关复制所期望的值。增加马尔科夫链的阶,对一阶自回归系数没有影响,很小的数值差异归因于低阶概率导出方法(附录 B,略)。

同样,利用链相关过程的普通形式也是低估二阶自相关系数(即表 3 中 0.046 对 0.113)。在这种情况下,马尔科夫链的阶自一阶增到二阶并考虑非恒等分布和自相关强度,全部贡献给二阶自相关系数的估计值,两

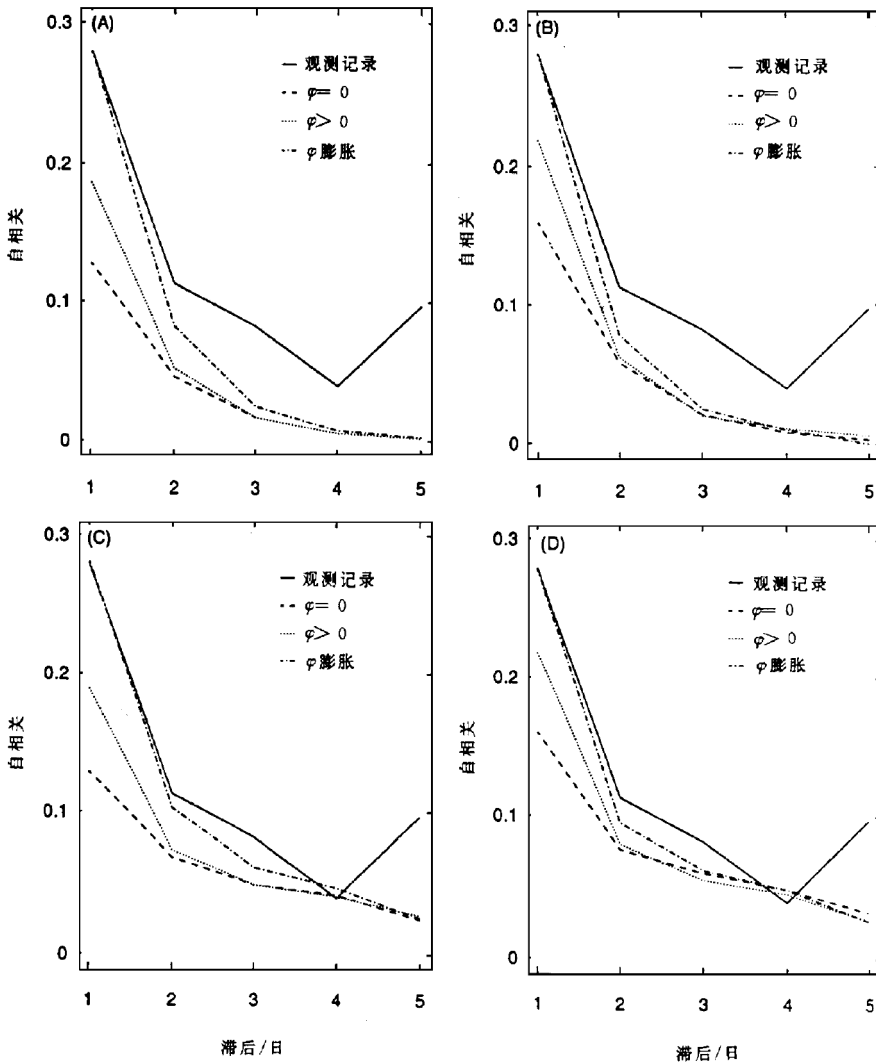


图 1 各种随机模式导出的加利福尼亚奇科 1 月份日降水量的自相关函数(曲线表明强度自相关的三种方案)以及样本自相关函数(实线)(a)降水出现一阶马尔科夫链和强度为恒等分布,(b)一阶与非恒等分布,(c)四阶与恒等分布,(d)四阶与非恒等分布

者组合产生的值为 0.078。当强度自相关也膨胀时,产生的最大值是 0.103(四阶链,恒等分布),仍略低于观测记录的值。认识到膨胀参数 ϕ 的这个方法不一定对重建滞后 $l \geq 2$ d 的自相关是重要的。Gregory 等(1993)研究英国降水时也发现了估算自相关时的这种不足。

图 1 表明这些随机模式子集(即马尔科夫链的阶限定一阶或四阶)的滞后到 $l=5$ d 的自相关函数,滞后 $l=3,4$ 与 5 d 是通过随机模拟估算的。对一阶链而言,显然是过低估计高阶的自相关(即滞后 $l \geq 2$),无论强度是恒等分布或非恒等分布以及自相关与否(图 1a,b)。另外,只要不认为观测的四阶以上的五阶自相关表现的增长是真实的,无论模式假设如何,四阶链的低估值显然是不多的(图 1c,d)。

3) 总降水

表 3 也包含随机模式多种形式估算的奇科 1 月份总降水的标准差,同样的数值也显示在图 2 中。对恒等分布强度无任何自相关的估计值是从(2)式任一阶马尔科夫链获得(利用雨日数的方差计算)。对具恒等分布自相关强度的一阶链,估计值是从(A1)与(9)式获得(对非恒等分布非相关强度,则根据(A1)与(12)式。在其他方面,估计值是基于随机模拟。

诚如 Katz 和 Parlange(1993)所预期的,链相关过程普通形式估算的标准差 68.7 mm(近似表达式(7)产生几乎相同的值 69.2 mm)与观测记录的标准差 88.6 mm 比较,是相对低的,或方差约 40%的超离差(表 3)。

具非恒等分布自相关强度的四阶链,最高的估计值是 80.6 mm(17%超离差)或还允许膨胀自相关,则为 83.9(10%超离差)。图 2 说明增加马尔科夫链的阶对估计值的标准差影响最大,强度自相关影响较小(粗略比较,如果让自相关膨胀),非恒等分布起作用最小(但要注意:膨胀的自相关能很好地代替非恒等分布强度的更复杂形式)。

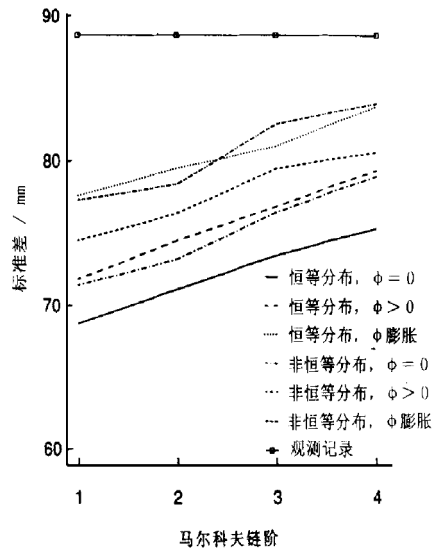


图 2 各种日降水量随机模式估算的加利福尼亚奇科 1 月份总降水量标准差与降水出现马尔科夫链模式阶的函数关系(曲线对应强度分布和自相关的六种方案,水平线为观测记录值)

无论怎样,显然通过利用更复杂的降水高频变化随机模式,即使不能消除超离差现象,也能使其程度大大地降低。我们注意到 Klugman 与 Klugmam(1981)也发现估算的俄勒冈州一个地点的季总降水的标准差对模式采用的形式敏感。问题仍然是,是否有比这里考虑的模式更复杂的模式能完全消除这种超离差。

3 讨论

已经证实,加利福尼亚奇科 1 月份总降水的超离差很大程度上可归因于随机模式对降水高频变化考虑不充分。日降水出现的高阶马尔科夫链模式完全消除雨日数超离差的来源。自相关和非恒等分布强度的修正量也有助于减少超离差。尽管适合于其他地方的日降水随机模式的形式可能与适合于奇科的模式有较大差异,但估算的月总降水的方差对模式所采用的形式照样敏感是意料之中的。(转 44 页)

(接 39 页)

能将月总降水的超离差进一步减少吗?通过日降水统计与大尺度大气环流的关系研究获得的结果,对这个问题作出了一些解释。Katz 和 Parlange(1993)对奇科 1 月份相同的日降水资料,给定大尺度大气环流指数,拟合了条件链相关过程最简单的形式。当这些条件模式组合成一个综合“次生”模式时,1 月份总降水的超离差减少 4%左右,还是小于本文所获得的减少量(即最好模式为 10%或 17%)。有趣的是次生模式完全消除了雨日数的超离差,与本文所获得的结果一致(Katz 和 Parlange,1996)。

本文有关估计潜在可预报性的结论仍有

待探索。虽然两种方法并不是等价的,但次生模式在某些方面类似一个能直接拟合资料的更复杂的模式(Katz 和 Parlange,1996)。进一步的研究应是寻找能统一这两种方法的途径。一种可能应包含所谓隐马尔科夫模式(hidden Markov models)及其推广(Guttorp,1995,第二章)。这些模式包含一种像大气环流指数一样但尚未观测过的隐藏状态。与隐藏状态结合的月总降水方差也许能解释为潜在可预报性的估算。

附录(略)

陈柏林译自 *Journal of Climate*, 1998, 11
(4), 591—601 马淑芬校