第 19 卷第 3 期　　南 京 气 象 学 院 学 报　　Vol. 19 No. 3
1996 年 9 月　　Journal of Nanjing Institute of Meteorology　　Sep. 1996

单站气温信息传递及其可预报性研究*

裕国　　况雪源
(南京气象学院应用气象学系, 南京　210044)　(广西壮族自治区气候中心, 南宁　530021)

摘　要　采用信息传递量作为度量气候可预报性的指标, 分别讨论了我国各大气候区的 10 个代表测站平均气温的可预报性。数值计算表明, 各种不同时间尺度(如候、旬、月、年) 信息传递量具有不同的衰减特征, 其差异反映出实际气候背景的差异, 信息传递量较好地代表了可预报性特征。由于实测的信息衰减与模式信息衰减具有一致性, 因此, 信息传递可作为检验模式可预报性的有益工具。

关键词　信息传递, 气候预报, 可预报性
分类号　P56.4

用何种指标描述气候的可预报性, 历来是学者们关注的焦点[1,2]。信噪比虽是常用的可预报性指标之一, 但也有一些缺陷。例如, 信噪比可给出不同初始距平或不同位置(地点) 的气候可预报性, 但却不能综合给出整个气候的可预报性[3~5]。因为不同地点的信噪比很难作简单的加权平均或内插, 尤其是当初始距平为零时, 信噪比并不能有效地度量气候可预报性。一般说来, 一个量要成为度量可预报性的有效指标, 应当是无论初始距平大小, 都可检测出它的可预报期限。因为实际大气中, 极端距平并不经常出现, 必须有一种客观方法, 按不同距平出现频率给以不同权重来考虑初始距平的总效应。本文在引入"信息传递"概念的基础上, 探讨实际气候可预报性的度量问题。以克服信噪比的某些不足。

1　信息传递及其计算

早在 70 年代, 文献[3]已经将信息与熵理论用于气象预告问题, 本文所引信息传递的概念, 等价于文献[4]中所定义的信息概念。

设某系统的熵为 $H(A) = -\sum_{i=1}^{n} P(A_i)\log P(A_i)$, 式中, A 表示互斥事件 $A_1, A_2, \cdots, A_n$ 的完备群, 对应概率分别为 $P(A_i)$, $i = 1, \cdots, n$。当状态出现概率相等时, 熵值达最大, 系统不确定程度最大。若 B 系统由状态 $B_1, B_2, \cdots, B_m$ 构成, 则两个系统同时出现的熵值应为 $H(AB) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i, B_j)\log P(A_i, B_j)$, 同样可写出条件熵的表达式

$$H(A \quad B) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i, B_j)\log\frac{P(A_i, B_j)}{P(B_j)}$$

所谓信息传递或信息, 是指系统的熵值与条件熵值之差。由于另一系统的存在给本系统带来的

* 国家自然科学基金资助项目
收稿日期: 1995- 09- 28; 改回日期: 1996- 03- 28

信息, 消除了若干不确定性, 于是有 $I(A,B) = H(A) - H(A\ \ B)$, 表示由 B 系统传递给 A 系统的信息, 亦可写出其逆形式 $I(B,A) = H(B) - H(B\ \ A)$ 。利用数学期望公式, 不难证明, $I(A,B)$ 的具体形式为

$$I(A,B) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i,B_j)\log[P(B_j/A_i)/P(B_j)]$$

式中, $P(B_j/A_i)$ 为在 A_i 状态下, B_j 出现的条件概率; $P(A_i,B_j)$ 为 A_i 与 B_j 的联合概率。Leung 等[4](1990) 利用简单的热量平衡模式研究可预报性, 从理论上导出关于初始距平场 T_0 到某一时间 t 的温度距平 T 之间的信息传递为

$$I(T,T_0) = -\frac{1}{2}\log[1 - \exp(-2t/\tau_0)] \tag{1}$$

其中, τ_0 为温度场球谐模振幅的张弛时间(就全球平均温度而言, τ_0 对孤立的空气柱为 58 天, 对位于海洋上空的气柱为 4.8 年)。上式表明, 全球温度距平的信息传递在 2～3 倍张弛时间内迅速衰减到零, 可见模式气候与衰减时间尺度相匹配。但是, 在实测气候资料序列中, 是否也有类似的性质, 其信息传递的衰减特征如何? 这正是本文研究的内容。

根据文献[3], 对平稳正态序列拟合自回归模式(记为 $AR(P)$, 利用线性算子的变换运算可以证明, 在 τ 时刻后的信息传递为

$$I(\tau) = -\frac{1}{2}\log(1 - r_1\alpha_1 - r_2\alpha_2 - \cdots - r_p\alpha_p)(1 + \psi_1^2 + \cdots + \psi_{\tau-1}^2) \tag{2}$$

式中, $\psi_1, \psi_2, \cdots, \psi_{\tau-1}$ 为传递系数, $r_1, r_2, \cdots, r_p$ 为各阶自相关系数。在特殊情形下, 当序列为马尔科夫链(即红噪声) 时, 其 τ 时刻后提供的信息量(或信息传递) 则为

$$I(\tau) = -\frac{1}{2}\log(1 - e^{-2\tau/T}) \tag{3}$$

显然, 当序列符合红噪声即 $AR(1)$ 时, 随着 τ 的增大, 信息量迅速减少。在给定最大步长 T 的情况下, 只有 $\tau \ll T$ 时才有较大的信息传递。

2　试验方案及资料预处理

选取北京、南京、沈阳、武汉、成都、南昌、广州、南宁、贵阳、兰州等 10 个站的逐月、逐旬、逐候、逐年平均气温资料, 除年平均气温取 100 年(1881～1980 年) 外, 其余资料年限统一取为 30 年(1951～1980 年) 。为了资料取舍方便, 已验证 1951～1980 年(30 年) 与 1951～1990 年(40 年) 资料统计特征无明显差异。

首先对月、旬、候资料作标准化处理, 消除年变化的影响, 其次将每 5 年、3 年、2 年的月、旬、候值构成一个时间序列。而对年平均气温序列则先消除趋势, 使序列平稳化。

其次, 应用 FPE 准则[5]对各站的所有资料序列选取适当的阶数 K, 拟合 AR 模式, 并使最终预测误差达到最小。一般说来, 气温大多服从正态, 因而对实测温度记录, 采用(3) 式是适当的。本文以各站实测的候、旬、月、年序列为对象, 根据序列的自相关结构, 试验其适合的自回归模型阶数, 并由(2) 或(3) 式计算相应的信息传递量。

(1) 候平均气温　表 1 以成都为例, 给出候序列试验结果(每 2 年构成一次试验序列, 其中 N= 144, 最大后延 m= N/12)。由表可见, 候序列的最优 AR 阶数 P= 1(约占 74%) , 它的信息传递平均达 τ= 4～5(候) , 可衰减至零(其中 τ= 3 时已很小)。表明在绝大多数年份其可预报限期为 4 候(即在 20 天以内) , 但在少数年份, 由于高阶自回归的影响, 信息衰减缓慢, 例如 P= 3 或 P= 5, 往往至 τ= 5(候) 仍有一定的信息量在传递。

表 **1** 成都候序列平均信息传递(单位: 比特)

Table 1 Averaged information transmission for interpentad temperature series from Chengdu bit

$AR(P)$	τ				
	1	2	3	4	5
$P=1$(74%)	0.129 0	0.004 0	0.000 2	0.000 0	0.000 0
$P=3$(13%)	0.104 5	0.039 5	0.034 0	0.033 5	0.031 0
$P=5\sim7$(13%)	0.147 0	0.071 0	0.062 5	0.061 5	0.060 5
平均	0.123 1	0.018 0	0.011 0	0.010 7	0.010 1

(2) 旬、月平均气温 表 2 所列为 5 个具有优势阶数($AR(1)$)的测站的旬际序列的平均信息传递量(其余测站类似, 从略), 抽样显示大多数测站旬际序列都符合马尔科夫链, 它们的信息传递平均约持续 $\tau=2$ 或 3(旬), 但也有的测站出现阶数为 2、3 或 4 的高阶 AR 情形(表中从略)。例如, 沈阳(40%的年份) $P=2$, 南宁、广州竟有 30% 的年份为 $P=3$ 或 2。这些试验序列的信息衰减, 因高阶 AR 序列中隐含准周期而十分缓慢。平均约持续 5～6 旬才衰减至零, 大多数测站 $P=2$ 或 3 持续至 5 旬即衰减为零。由此可见, 在不少年份, 因序列中隐含高阶 AR(等价于准周期序列)而使可预报极限期加长。表 3 为 6 站月际序列的优势 AR 序列情形下的信息传递平均量。

表 **2** **5** 个代表测站旬际序列 $\boldsymbol{AR(1)}$ 信息传递平均量

Table 2 Averaged information transmission for interdecadal temperature series $AR(1)$ of five of the selected stations bit

	τ				
	1	2	3	4	5
北京(0.7)	0.012 3	0.000 1	0.000 0	0.000 0	0.000 0
南京(0.9)	0.017 4	0.002 3	0.002 2	0.000 1	0.000 0
成都(0.7)	0.030 6	0.002 6	0.000 3	0.000 0	0.000 0
南昌(0.8)	0.022 1	0.001 3	0.000 0	0.000 0	0.000 0
贵阳(0.7)	0.011 6	0.000 1	0.000 0	0.000 0	0.000 0

说明: 括号内为该序列出现频数百分率

计算表明, 对绝大多数测站(除贵阳、武汉外)来说, 月序列的最优阶 $P=1$, 一般都有≥50% 的年份, 其信息衰减约持续 $\tau=1\sim4$(月), 平均 2～3(月), 总的说, 高阶 AR 所占比例不大(特殊的只有武汉、贵阳), 一般约持续 $\tau=4$(月)。由上可见, 月序列信息传递持续时间最可能在 $\tau=2$(月)。换言之, 月序列的可预报期限也仅为两个时间步长。只在个别的站点或年份才有可能符合高阶而具有较多的信息传递。

(3) 年平均气温 从表 4 可见, 除南昌 $P=1$ 信息衰减较快, 其他站皆为高阶 $AR(P)$, 成都、兰州站尤为显著, 且这两站点在 $\tau=1$ 时信息传递量仍较大, 而其他各站多属 $AR(2)$, 信息传递在 $\tau=1$ 时就较小, 以后衰减很缓慢。可见, 在年际序列中很可能存在准周期的影响, 从而不能简单以 $AR(1)$ 拟合。

表 3　6 个代表测站月际序列($AR(1)$)信息传递平均量

Table 3　Averaged information transmission for intermonthly temperature series of six of the selected stations　　bit

	τ				
	1	2	3	4	5
北京(100%)	0.019 3	0.001 2	0.000 0	0.000 0	0.000 0
南京(83%)	0.051 8	0.008 2	0.001 0	0.000 0	0.000 0
沈阳(100%)	0.010 7	0.000 0	0.000 0	0.000 0	0.000 0
广州(100%)	0.014 0	0.001 0	0.000 0	0.000 0	0.000 0
南宁(100%)	0.064 0	0.000 7	0.000 0	0.000 0	0.000 0
兰州(100%)	0.071 2	0.032 0	0.000 8	0.000 2	0.000 0

说明: 括号内为该阶数的出现频数百分比

表 4　各站百年序列信息传递量

Table 4　Averaged information transmission of intercenturial temperature series for each of the study stations　　bit

τ	1	2	3	4	5
北京(P= 2)	0.064	0.031	0.029	0.028	0.028
南京(P= 2)	0.027	0.018	0.018	0.018	0.016
沈阳(P= 2)	0.048	0.037	0.036	0.036	0.036
成都(P= 4)	0.151	0.125	0.124	0.123	0.121
南昌(P= 1)	0.041	0.002	0.000	0.000	0.000
南宁(P= 2)	0.069	0.039	0.038	0.036	0.036
贵阳(P= 2)	0.081	0.043	0.041	0.039	0.039
兰州(P= 10)	0.137	0.101	0.099	0.099	0.099

3　信息传递特征的年际变动

本文在资料预处理过程中, 采用分段子序列多层抽样方法, 使各子序列不但具有大样本, 而且均满足平稳性假设。从这个意义上说, 增加了同类序列的可比较性(见表 1～3) 。

大多年代的子序列, 符合 $AR(1)$, 但仍有少数年代子序列属高阶自回归。以成都为例, 在月值序列中: 1966～1970 年序列为 $AR(4)$, 1971～1975 年为 $AR(2)$; 旬值序列中: 1963～1965 年为 $AR(9)$, 1967～1971 年为 $AR(6)$, 1978～1980 年为 $AR(2)$; 候值序列中: 1951～1952 年为 $AR(5)$, 1957～1958 年为 $AR(3)$, 1967～1968 年为 $AR(7)$, 1971～1972 年为 $AR(3)$ 。其余时间段皆为 $AR(1)$ 。这是否由于抽样误差或气候背景的影响, 将在后面加以讨论。

4　不同时间尺度信息传递特征的平均差异

从表 1～4 可见, 当自回归阶数为 1 时, 月值序列的信息传递量衰减很快, 各站虽有不同, 但基本在 τ= 2～3 时衰减至零。旬值序列比月值序列信息传递量衰减慢, 虽也约在 τ= 2～3 时衰减至零, 但相同 τ 值时的信息传递量明显比月值序列大。候值序列衰减减慢更为明显, 约在 τ

= 4～5 时方衰减至零。还可看到, 相同阶数的高阶自回归序列也有此特性。可见随时间尺度的增加, 单站同阶同年代的信息传递量衰减总是加快, 亦即说明短时间尺度的可预报性一般总比长时间尺度的可预报极限期大些。

此外, 将各站不同年代的信息传递量及自相关函数分别求其平均衰减状况(图略) , 结果发现, 在年际序列中, AR(1) 的信息衰减最快, 其相应的自相关函数围绕时间后延轴呈波动衰减至零, 也较迅速。但是, AR(2) 的信息传递衰减却很缓慢且趋微弱, 而相应的自相关函数也有类似规律。不过, 在候、旬、月际序列中, AR 的阶数与自相关函数的关系并不太明显。

5 结 语

对各大气候区 10 个代表测站的平均气温信息传递特征的分析结果表明:

(1) 候、旬、月际气温序列, 符合 AR(1) 的机会为最多, 而其信息传递衰减与马尔科夫链的理论基本吻合, 绝大多数都在 τ= 2～3 时即衰减至零, 仅个别序列(如候际序列) 稍延迟到 τ= 5 衰减至零, 这表明了大致的可预报期限。

(2) 信息传递确实对气候可预报性有指示意义。上述结果与 *Leung* 等人对模式气候的验证结果基本一致, 说明了信息传递无论对模式验证或资料检测都有重要价值。

(3) 信息传递随时间尺度不同而有不同的衰减特征, 尤其是单站同阶同年代序列随时间尺度的增大, 其信息传递量衰减加快。这也在一定程度上表明了候、旬、月、年际序列的可预报期限的大致尺度。

由于年、月、旬、候际序列中存在着一定比例的高阶自回归特征, 并不能完全以简单 AR(1) 拟合, 因此, 可预报期限实际还存在某些年际变动或阶段性, 即不确定性。因为它们实际上是与气候的年际振动背景有关的。为了考察这一现象, 我们同样对相应年份的降水作自回归拟合的尝试。结果发现, 在高阶自回归的年份, 降水量的起伏波动很大, 而在 AR(1) 年代段则相反, 降水相对较平稳且离差较小。这从另一侧面反映出气候的年际振动会影响各年份内部的时间序列结构。此外由于模式和实际资料所计算的信息传递特征在一定的程度上具有一致性, 因而, 计算模式的信息传递特征与实际信息传递特征的吻合程度, 将不失为一种评价模式优劣的方法。

参 考 文 献

1 *Leith C B. The standard error of time-average estimates of climatic means. JAM*, 1973, 12: 1066～1075

2 *Hayashi Y. Confidence intervalues of a climatic signal. JAS*, 1982, 39: 1895～1905

3 张学文 . 气象预告问题的信息分析 . 北京: 科学出版社, 1981. 1～48

4 *Leung L Y, North G R. Information theory and climate prediction. J Climate*, 1990, 3: 5～14

5 么枕生, 丁裕国 . 气候统计 . 北京: 气象出版社, 1990. 1～35

6 黄嘉佑 . 长期天气过程的可预报性研究的进展 . 气象科技, 1992, (3) : 30～37

STUDY OF TEMPERATURE INFORMATION TRANSMISSION AND ITS PREDICTABILITY FOR SINGLE STATION OF CHINA

Ding Yuguo
(*Department of Applied Meteorology, NIM, Nanjing* 210044)
Kuang Xueyuan
(*Guangxi Climatic Research Center, Nanning* 530021)

Abstract The predictability of mean temperature is investigated for ten stations located, separately, in all climatic regions of the country in terms of information transmission as the measure of climatic predictability. Evidence suggests that the transmission at a range of scales (pentad, decade, month and year) displays attenuation features of their own in relation to the climatic background. As such, the temperature information transmission represents better the predictability. The consistent attenuation in the measured and model-yielded temperature information allows to use the transmission as a useful means for testing the predictability dealt with by a model.

Keywords information transmission, climatic prediction, predictability