

# 多模态模型及其在气候分析中的应用<sup>1)</sup>

王 蕾 康杜鹃<sup>2)</sup> 李晓东<sup>3)</sup>

(北京大学地球物理系,北京,100871)

**摘 要** 对气候要素的概率分布进行了讨论,根据这些概率分布的特点提出了多模态模型,并用非线性拟合的方法进行了验证。对赤道平流层纬向风、ENSO指数(ESI)序列、冰岛 Reykjavik(21.90°W,64.13°N)气压和瑞典温度、北半球树木年轮等数据的分析表明:气候系统中多模态现象具有普遍性。

**关键词** 概率分布;多模态;平衡态;非线性拟合

中图分类号 P468.01

## 0 引言

传统上,人们用气候要素的统计平均值和相对于平均值的偏差来描述气候及气候变化,而且也尝试用线性回归或ARMA模型等方法来对气候要素进行预报<sup>1,2)</sup>。然而这无疑相当于给气候系统强加了这样的前提——气候要素服从正态分布<sup>3)</sup>,围绕着一个均值变化,并且这种变化是平稳的。但是,由于气候系统是高度非线性的复杂动力系统,所以实际情况并非如此。可以轻易地举出反面的例子:热带平流层的准两年振荡(QBO)、季风等等,理论上也是如此,由大气运动方程简化而来的Lorenz系统就是一个极好的例证。

本文在讨论有关正态分布的数学物理背景的基础上,对一些具有代表性的各种时间尺度的气候要素进行统计分析,发现许多典型的气候要素都不服从正态分布,而具有多峰结构。

## 1 正态分布的数学和物理背景

统计学方法在物理学中的最初尝试可以追溯到Maxwell和Boltzman的气体理论,但仅用来描述各个可能的状态以及这些状态存在的可能性——而不是作为系统时间演变的一个内在属性。因此,实际上必须把Einstein对布朗运动的解释视为自然现象随机模型的开端<sup>4)</sup>。于是,正态分布在物理学领域捷足先登了。此后,由Langevin引入的随机微分方程更直接地描述了这一现象<sup>4)</sup>。在布朗运动这个例子中,“单个分子与花粉微粒的相互作用是同一类型的”,而气候系统的各种外强迫的机制却相去甚远,这似乎阻碍了正态分布的应用,但是,我们有“中心极限定理”<sup>5)</sup>。中心极限定理表明,无论各个独立的随机变量 $X_k(k=1,2,\dots,n)$ 服从什么分布,

1) 国家自然科学基金资助项目(G1999043400)

2) 现在工作单位:中国科学院大气物理所,北京,100029

3) 通讯联系人

只要满足定理要求的条件,那么当  $n$  很大时它们的和就近似地服从正态分布。这就是正态随机变量在概率论中占有最重要地位的一个基本原因。定理的物理背景是,如果一个系统受到大量“影响一致小”且相互独立的随机因素的作用,那么系统随时间的演变将是正态随机过程<sup>[3]</sup>。在气候系统中,定理要求的条件——Lindeberg 条件——往往很容易被满足。于是,对于气候系统这样复杂的对象,也有了强有力的统计学工具。正是因为正态分布具有这样重要的地位,当人们研究一个连续的总体时,往往首先考虑它是否服从正态分布,以及它的演变过程是否为正态随机过程。但往往是,正态分布被滥用了,在后面的实例中我们可以看到,众多气候要素的概率分布根本就通不过正态检验。

对于动力系统而言,其系统变量的概率密度也称作自然测度(Natural Measure)或不变测度(Invariant Measure)。只有在保证资料的遍历性条件下才能讨论不变测度,不变测度是对系统进行统计分析的基础。同 Lyapunov 指数、分维谱一样,不变测度也是动力系统最重要的全局不变量。一个动力系统只能对应一个唯一确定的不变测度。Gauss 随机过程的不变测度一定是正态分布<sup>[4]</sup>,其逆否命题也肯定成立——如果一个动力系统的不变测度不是正态的,则此系统的内部过程一定不是正态随机的。

气象气候学家早就注意到某些气象气候要素的概率分布完全不同于正态分布<sup>①</sup>,例如平流层准两年振荡(QBO)<sup>[5]</sup>,因此上述命题的一个自然推论是 QBO 信号不是源于一个正态随机系统。关于 QBO 的被普遍接受的动力学解释是由 Lindzen 和 Holton 提出的波流相互作用模型<sup>[6]</sup>,这个非线性不稳定性模型在 Plumb 和 McEwan<sup>[7]</sup>的实验室相似性试验中已经得到证实。Charney 和 Drazin<sup>[8]</sup>已证明,只要平均纬向气流为西风且小于某一临界值  $U_c$ (称为“Rossby 临界速度”),行星波就能垂直传播到平流层之中。这说明平流层纬向风处于东风位相和西风位相时,系统处于两种截然不同的动力过程。这就必然要求重新考虑用两个正态随机过程来分析 QBO,以区别对待上述两种动力过程。实际上,可以容易地举出具有双模态特征的简单的物理系统,例如 W 型的势阱中单个质点在随机外强迫作用下的运动。因此,本文提出基于不变测度的多模态模型:系统变量围绕着多个最可几状态按正态分布变化,并且围绕每个最可几状态的变化是平稳的。这些最可几状态就是热力学和统计物理中所谓的平衡态,由于气候系统是一个开放系统,这种平衡是动态的,在统计性质上就表现为围绕平衡态的某种概率分布——在许多情况下是正态分布。

## 2 多模态模型

由多模态模型思路,气候要素可能表现为围绕多个最可几状态变化,在不变测度中的反映就是总的概率分布可视为多个正态分布的线性迭加。因此本文的拟合模型为

$$f(x) = \sum_{i=1}^n \frac{\alpha_i}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right], \quad (1)$$

其中参数  $\alpha_i, \mu_i, \sigma_i$  ( $\sigma_i > 0, i = 1, 2, \dots, n$ ) 分别表示第  $i$  个正态分布的权重、期望和标准差,并且有约束关系  $\sum_{i=1}^n \alpha_i = 1$ 。在最小二乘意义下对概率分布进行拟合,模态数  $n$  可以根据  $\chi^2$  检

验来确定,作者采用了与此类似的最小描述长度( Minimum Description Length, 简称 MDL )<sup>[9]</sup>来进行选择。这实际上是一个非线性拟合问题。本文的拟合程序采用了 Levenbery-Marguardt 迭代算法<sup>[10]</sup>,由于算法的初值敏感性,应根据分布曲线的形状对初始参数进行适当的控制并采取加入随机扰动的方法以期尽可能逼近全局最优。

本文用于多模态拟合的气候资料包括:平流层纬向风、ENSO 指数(ESI)序列、冰岛 Reykjavik 气压和瑞典温度、北半球树木年轮资料。

在拟合前首先进行偏度和峰度检验,以判断该序列是否符合正态分布。随机变量  $X$  的偏度和峰度分别指  $X$  的标准化变量  $\frac{x-\mu}{\sigma}$  的三阶中心矩  $\nu_3$  和四阶中心矩  $\nu_4$ 。设  $x_1, x_2, \dots, x_m$  是来自总体  $X$  的样本,  $g_1, g_2$  分别为样本偏度和样本峰度。当随机变量  $X$  服从正态分布且样本容量  $m$  足够大时,统计量  $U_1 = g_1/\sqrt{6/m}, U_2 = (g_2 - 3)/\sqrt{24/m}$  都近似地服从  $N(0, 1)$ 。在给定的信度  $\alpha$  下,正态假设的拒绝域为  $|U_1| \geq k_1$  或  $|U_2| \geq k_2$ 。一般地取  $\alpha = 0.1$ , 有  $k_1 = k_2 = 1.96$ 。

如果序列通不过正态性检验,就进行多模态拟合:从两个模态开始依次增加模态数,每次拟合后计算相应的最小描述长度(MDL),直到其最小描述长度小于低一阶拟合的 MDL 值——认为此时模态数的增加对于数据拟合的贡献已不那么明显。

### 3 多模态模型的气候分析应用

资料的遍历性问题是拟合时必须要考虑的。在以下的应用例子中,尽管可能已经保证了资料的遍历性,但是部分资料仍有可能得不到光滑的频率分布。在这种情况下,为了得到更可靠的拟合结果,作者对原始序列进行滑动平均处理以消除噪声。

#### 3.1 准两年振荡(QBO)

在平流层大气环流的变化中存在明显的准两年振荡(QBO)<sup>[5]</sup>,主要表现为热带平流层纬向风和温度场,其周期是不规则的,一般稍长于两年。QBO 是全球气候的年际变化中的一个强信号,其双模态特征是非常显著的。如前所述,QBO 的动力机制已经有了较成熟的解释<sup>[6-8]</sup>。

本文使用的 QBO 原始资料是 1949-01-01—1998-12-31 共 18 628 天的赤道东经  $115^\circ 30$  hpa 等压面纬向风速序列。该时间序列标准差为 11.09,显然这与实际情况是不符的——该值明显大于平均的平流层纬向风逐日变化幅度。这也说明用单模态正态分布来描述平流层纬向风是不合理的。对平流层纬向风偏度、峰度检验结果表明偏度( $U_1$ )与峰度( $U_2$ )分别为  $-5.59$  和  $-23.13$ ,在显著水平  $\alpha = 0.1$  下,可以认为该分布并非正态分布。

采用第 2 部分所表述的模型来对该分布曲线进行拟合,其中模态数  $n$  取为 2,表示总分布由两个正态分布线性迭加而成。拟合结果见表 1,其中  $Q$  表示拟合的残差平方和, $r$  表示拟合结果与原概率分布的相关系数。

与表 1 相对应,图 1 给出了拟合结果与实际分布情况的对比,从图中可以清楚地看到拟合曲线的双峰结构,而 2 个正态函数的位置、形状及各自对总体分布的贡献也可以从图上直观地获得。直方图表示实际的概率分布情况,粗实线为拟合得到的概率线,细实线则分别表示了构成总体分布的 2 个正态函数(下同)。由此可见,热带平流层的准两年振荡明显地不同于正态

表 1 QBO 的双模态拟合结果

Table 1 The fitting result of the bimodal of QBO

模态	$\alpha_i$	$\mu_i$	$\sigma_i$	$Q$	$r$
I	0.329	7.358	5.372	4.371e-5	0.9977
II	0.678	-9.806	9.008		

随机过程,在对它起主要影响的内部物理过程有 2 种,拟合得到的 2 个正态函数恰好可以反映这 2 种过程的特征。2 种过程的自身属性决定了这 2 个正态函数的期望和方差,而 2 个正态函数权重代表它们的影响程度——与各个模态发生的频率有关。经过最小描述长度(MDL)的检验,证明采用两个模态进行拟合相对其他模态数(1 个、3 个或更多)是最优的,模态数为 1、2、3 时相应的 MDL 分别为 -320.6、-414.1、-413.1。

### 3.2 厄尔尼诺-南方涛动(ENSO)

南方涛动(SO)是指热带东太平洋地区和热带印度洋地区气压场反向变化的跷跷板现象,一般采用南方涛动指数(SOI)<sup>[11]</sup>来对其进行描述。许多研究都表明,南方涛动与赤道东太平洋海表水温异常事件(El Nino)之间有非常强的负相关,其相关系数在 -0.57 ~ -0.75 之间,信度达到 99.9%。通常认为 ENSO 包含 3 种比较明显的状态<sup>[12]</sup>,即暖状态(El Nino 和负 SOI 指数)、冷状态(La Nina 和正 SOI 指数)以及正常态(或过渡态)。这些状态会循环出现,是 ENSO 循环的振荡周期在 3 ~ 7 (年)的范围内变化。El Nino 和 SO 是耦合的海洋—大气系统中一个全球尺度振荡(ENSO)的两个方面<sup>[2]</sup>。类似于王绍武等<sup>[12]</sup>的处理,作者将 Nino3 区的海表温度和 SOI 标准化后相减得到 ENSO 指数 ESI(ENSO Index):

$$ESI = SST_{\text{norm}} - SOI_{\text{norm}}, \quad (2)$$

以期能够客观地反映出 ENSO 的模态特征。

本文使用的原始资料为从 1876-01 到 1997-12 共 1464 个月的 SOI 序列和 Nino3 区的海表温度<sup>[12]</sup>。王绍武等<sup>[12]</sup>指出,季分辨率的 ENSO 指数可以较好地确认近百年来 ENSO 事件及其强度;另一方面,作季平均还可以过滤掉局地因素产生的小波动。因此这里对原始序列作了 3 点滑动平均。

对该分布的正态性检验表明,ESI 偏度、峰度检验结果分别为 7.49、2.57,均未通过显著水

表 2 ESI 的三模态拟合结果

Table 2 The fitting result of the trimodal of ESI

模态	$\alpha_i$	$\mu_i$	$\sigma_i$	$Q$	$r$
I	0.094	-1.401	0.383	4.742e-5	0.9993
II	0.515	-0.371	0.531		
III	0.385	0.769	0.731		

万方数据

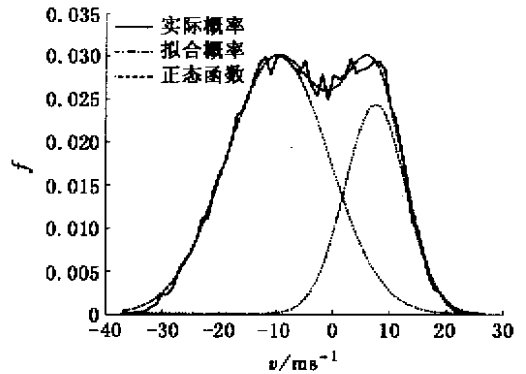


图 1 平流层纬向风速的实际分布与拟合分布曲线

Fig.1 The actual frequency distribution and the fitting curve of stratospheric zonal wind

平  $\alpha = 0.1$  的正态分布检验。经过最小描述长度(MDL)的检验,证明采用 3 个模态进行拟合相对其他模态数是最优的,前 4 个模态数对应的 MDL 依次为 -32.75、-184.76、-189.17、-183.93。由此可见,ENSO 循环的确包含 3 种明显不同的模态,拟合结果见表 2 和图 2。

### 3.3 冰岛 Reykjavik (21.90°W64.13°N) 气压和瑞典温度

前两个例子都是主要表现年际尺度气候变化特征的, 本节两个例子是日分辨率的资料。资料是冰岛 Reykjavik 气压和瑞典温度的逐日记录<sup>[13]</sup>。气压序列开始于 1822-03-01, 到 1999-10-04 结束; 温度序列开始于 1772-03-01, 到 1998-12-31 结束。这 2 个序列的偏度、峰度均未通过显著水平  $\alpha = 0.1$  的正态分布检验。

表 3、表 4 和图 3、图 4 给出了拟合结果与实际分布情况的对比。经过最小描述长度 (MDL) 的检验, 证明采用两个模态进行拟合相对其他模态数是最优的。气压序列的 MDL 依次为 -230.0, -492.2, -478.8, 温度序列的 MDL 依次为 -427.9, -1255.7, -1250.2。

表 3 Reykjavik 气压的双模态拟合结果

Table 3 The fitting result of the bimodal of Reykjavik surface pressure

模态	$\alpha_i$	$\mu_i$	$\sigma_i$	$Q$	$r$
I	0.499 3	9 986	148	1.200e-5	0.999 8
II	0.499 6	10 128	99		

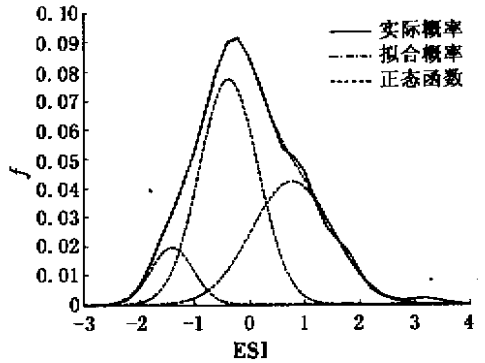


图 2 ESI 的实际分布与拟合分布曲线

Fig.2 The actual frequency distribution and the fitting curve of ESI

表 4 瑞典温度的双模态拟合结果

Table 4 The fitting result of the bimodal of Swedish temperature

模态	$\alpha_i$	$\mu_i$	$\sigma_i$	$Q$	$r$
I	0.301	14.836	2.393	1.774e-5	0.998 8
II	0.696	6.828	4.352		

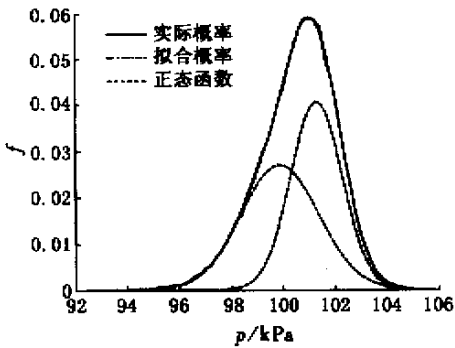


图 3 Reykjavik 气压的实际分布与拟合分布曲线

Fig.3 The actual frequency distribution and the fitting curve of Reykjavik surface pressure

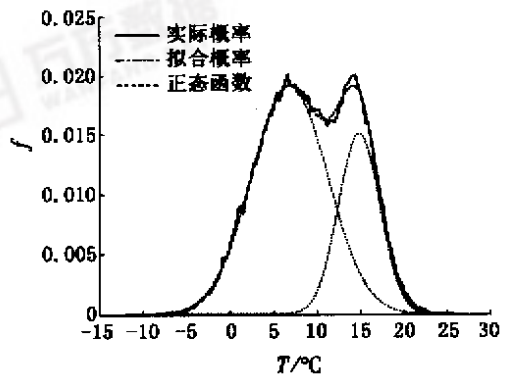


图 4 瑞典温度的实际分布与拟合分布曲线

Fig.4 The actual frequency distribution and the fitting curve of Swedish temperature

### 3.4 北半球树木年轮

对于百年到千年尺度的气候变化的细节和机制仍有许多问题有待进一步研究, 但就近 600 年北半球树木年轮资料<sup>[14]</sup>而言, 也具有明显的多模态特征。对其偏度峰度检验表明, 该序



列不遵从正态分布。对其进行多模态拟合,结果如图 5 和表 5 所示。由图 5 可见,由 MDL 确定的模态数为 3 的多模态模型可以很好地逼近树木年轮序列的实际分布。

表 5 北半球树木年轮的双模态拟合结果

Table 5 The fitting result of the bimodal of tree ring

模态	$\alpha_i$	$\mu_i$	$\sigma_i$	$Q$	$r$
I	0.483	-0.278	0.207		
II	0.435	-0.132	0.111	$1.721e-4$	0.998 8
III	0.077	0.0881	0.049		

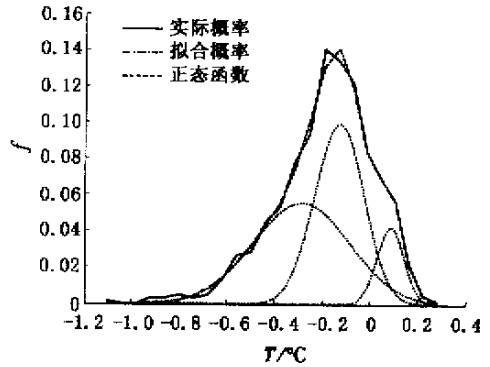


图 5 北半球树木年轮的实际分布与拟合分布曲线

Fig.5 The actual frequency distribution and the fitting curve of tree ring

## 4 结论与讨论

1) 气候系统的各种外强迫机制不规则,例如太阳辐射、地球轨道参数、温室效应等等;海气相互作用等内部过程高度非线性、极端复杂。因此用单个正态分布来描述气候系统是有局限性的。

2) 对气候系统中多个时间尺度资料(平流层纬向风、ENSO 指数(ESI)序列、冰岛 Reykjavik 气压和瑞典温度以及北半球树木年轮等)的分析表明:多模态现象在气候系统中具有普遍性。

3) 系统的模态数决定于其物理机制,统计检验只是系统内部机制在信号上的反映。尽管每个峰态是否是严格的正态分布还有待于进一步确认,但气候要素的概率密度分布明显具有多峰态结构。本文的拟合只是在遍历性假设下的一种逼近方法,由于气候资料在遍历性和精度等方面还有待完善,因此本文的结果可以进一步精细化。

4) 气候多模态模型在气候系统研究中的应用前景是广阔的,例如气候突变的检测、气候模式的参数化、气候预报等等。

## 参 考 文 献

- 1 李晓东. 气候物理学引论. 北京:气象出版社,1997
- 2 Jose P Peixoto, Abraham H Oort. 吴国雄,刘辉译. 气候物理学. 北京:气象出版社,1995
- 3 中国大百科全书出版社编辑部. 中国大百科全书(数学卷). 北京:中国大百科全书出版社,1980
- 4 Gardiner C W. Handbook of Stochastic Methods, for Physics, Chemistry and the Natural Sciences(2<sup>nd</sup> ed). Berlin: Springer-Verlag, 1985
- 5 Reed R J, Campbell W J, et al. Evidence of a Downward Propagating Annual Wind Reversal in the Equatorial Stratosphere. J Geophys Res, 1961, 66: 813 ~ 818
- 6 Lindzen R S, Holton, et al. A Theory of the Quasi-biennial Oscillation. J Atmos Sci, 1968, 25: 1095 ~ 1107
- 7 Plumb R A, McEwan A D, et al. The Instability of a Forced Standing Wave in a Viscous Stratified Fluid: a Laboratory Analogue of the Quasi-biennial Oscillation. J Atmos Sci, 1978, 35: 7378 ~ 7384
- 8 Charney J G, Drazin D G, et al. Propagation of Planetary Scale Disturbances from the Lower into the Upper Atmosphere. J Geophys Res, 1961, 66: 83 ~ 109
- 9 Kevin Judd, Anstijn Mees. On Selecting Models for Nonlinear Time Series. Physica D, 1995, 82: 426 ~ 444

- 10 Press William H. Numerical Recipes—The Art of Scientific Computing ,2nd ed. New York :Cambridge University Press ,1992
- 11 Walker G T ,Bliss E W. World Weather V. Mem R Meteor ,1932 ,Soc 4 53 ~ 84
- 12 王绍武 龚道溢. 近百年来的 ENSO 事件及其强度. 气象 ,1999 25( 1 ) 9 ~ 14
- 13 Moberg A Bergstrom H. Homogenization of Swedish Temperature Data ,Part III —the Long Term Temperature Records from Stockholm and Uppsala. International Journal of Climatology ,1997 ,17 667 ~ 699
- 14 Fritts H C. Reconstructing Large Scale Climatic Patterns from Tree-ring Data. Tucson ,Arizona :The University of Arizona Press ,1991

## Multimode and Its Application in Climate Analysis

WANG Lei KANG Dujuan LI Xiaodong

( Department of Geophysics ,Peking University ,Beijing ,100871 )

**Abstract** The probability distributions of climatic elements are discussed. Based on the characters of them ,a multimode model is introduced and verified through the nonlinear fitting. The climate features of quasi-biennial oscillation of the stratospheric zonal wind over equator ,the ENSO Index ,the surface pressure at Reykjavik( 21.90°W 64.13°N )and Swedish temperature ,and the Northern Hemisphere tree ring all show that the multimode are universal in the climatic system.

**Key words** probability distribution ; multimode ; equilibrium state ; nonlinear fitting

\* \* \* \* \*

校内要闻

### 北京射频超导研究中心在北京大学成立

2001年10月18日,在北京大学交流中心召开了由北京大学和中科院高能物理研究所联合组建的“北京射频超导研究中心”成立大会。许智宏校长、中科院白春礼副院长、国家基金委陈佳洱主任、高能物理研究所陈和生所长以及教育部科学技术司、科技部基础研究所等有关单位的领导也出席了成立大会。许智宏校长和白春礼副院长为中心揭牌。

许校长和中科院白春礼副院长在致词中指出,中科院高能所和北京大学携手共建“北京射频超导中心”是非常必要的,这种强强联合有利于进一步加强学术交流、有利于合作开展研究、有利于优势互补提高双方已有资源的利用效率、有利于更多年轻学者和学生参与国家重大科学研究项目。

高能所陈和生所长、方守贤院士、谢家荣院士、陈佳洱主任都在会上发了言。北京大学赵夔教授和高能所张闯副所长还在会上做了精彩的学术报告,介绍了双方目前研究工作的进展情况。会后,与会领导和专家饶有兴趣地参观了我校射频超导实验室,并在参观过程中对相互感兴趣的问题进行了切磋。

(摘自《北京大学校刊》网络版)