0577-6619/2010/68(4)-0501-13 *Acta Meteorologica Sinica* 气象学报

基于矩阵理论的全球气温的时空关联性研究*

支 蓉[1] 封国林[1,2] 龚志强[1] 周 磊[1]
ZHI Rong[1] FENG Guolin[1,2] GONG Zhiqiang[1] ZHOU Lei[1]

1. 扬州大学物理科学与技术学院,扬州,225002
2. 国家气候中心,中国气象局气候研究开放实验室,北京,100081
1. *Department of Physics Yangzhou University*, *Yangzhou* 225002, *China*
2. *Laboratory for Climate Studies*, *National Climate Center*, *China Meteorological Administration*, *Beijing* 100081, *China*
2008-03-02 收稿,2008-04-18 改回.

Zhi Rong, Feng Guolin, Gong Zhiqiang, Zhou Lei. 2010. Temporal and spatial correlation analysis of global temperature data based on the correlation matrix theory. *Acta Meteorologica Sinica*, 68(4):501-513

Abstract The temperature correlation matrix C and random correlation matrix R are constructed based on the NCEP/NCAR temperature reanalysis data and random series with the characteristics of these two matrixes analyzed in this paper. The results show that there are genuine correlations as well as the correlation 'noise' existing in the temperature correlation matrixes. These genuine correlations can be divided to two parts, one is the correlation between the nearest and next-nearest points, called the short distance correlation (SC); the other is the long distance correlation (LC), i. e. the correlation between the El Niño area and other remote areas such as the warm pools. For the different scales, these two kinds of correlations show different features. On the 1 - 10 d scale, the SC is more important than the LC, while on the 15 d and larger scales, the SC and LC both play an important role in the temperature correlation matrixes. Most correlation information is contained in several eigenvectors with larger eigenvalues, and the projection of the global temperature series on these eigenvectors will show, in some cases, the whole characteristics of global temperature changes. Besides, the temperature correlations have significant temporal and spatial variabilities: for example, the correlation is better during 1950 - 1956, 1972 - 1977 and 1996 - 2000 than during 1978 - 1982 and 1991 - 1996. On the 1 d scale, the correlation shows however a good latitudinally symmetric spatial distribution, but it is relatively worse on the 15 d scale owing to the ocean-land difference between the Northern and Southern Hemisphere. In contrast, on the 15 d scale, the correlation shows a longitudinally symmetric distribution.
Key words Matrix theory, Correlation coefficient, Eigenvalue, Eigenvector, Spatial distribution, Temperature

摘 要 采用1948—2005年NCEP/NCAR全球气温和随机序列分别构建气温关联矩阵和随机关联矩阵并进行比较,发现气温关联矩阵既存在关联"噪声",又存在真实关联;格点气温序列之间既存在临近格点间的短程关联,又存在如厄尔尼诺、三大洋暖池等区域格点间的远程关联。不同时间尺度,这两种关联的表现各有差异,1—10天等小尺度情况下,格点间的关联以短程关联为主;15天及以上尺度短程关联比例显著下降,长程关联有所上升。对关联矩阵和随机关联矩阵求解本征值和本征矢量发现,格点气温序列之间的关联信息主要包含在几个较大本征值所对应的本征矢量中,且格点气温序列在这些本征矢量上的投影能够在一定程度上体现全球气温变化的整体特征。此外,格点气温序列的关联性在时间和空间上均存在显著的变异性,主要表现为1950—1956年、1972—1977年和1996—2000年3个时段相对较高,而1978—1982年和1991—1996年则均值相对较低,存在显著的准10—20年周期振荡;关联系数的空间分布在上述两类尺度下均表现为沿纬向呈准带状分布,但1天

* 资助课题:国家重点发展基础研究项目(2006CB400503),国家自然科学基金项目(40675044),国家科技支撑计划(2007BAC03A01和2007BAC03B01)。
通讯作者简介:龚志强,主要从事气候突变和极端事件的检测与分析及气候系统复杂网络等方面的研究。E-mail: gzq0929@126.com

尺度的纬向平均具有较好的对称性,由于海陆差异的原因,15 天尺度的纬向对称性相对较差,而 15 天尺度的经向平均具有一定的准对称性。

关键词 矩阵理论,关联系数,本征值,本征矢量,空间分布,温度

中图法分类号 P468

1 引 言

气候系统是由大气圈、水圈、陆地表面、冰雪覆盖层和生物圈构成,每个圈层各自特殊的物理属性及复杂的尺度结构和物理结构,使气候系统成为一个高度非线性、多层次、强迫耗散的复杂系统。太阳辐射和大气环流作用下,气候系统内部产生一系列复杂的物理过程和非线性相互作用,系统各个组成部分之间通过物质和能量的交换等紧密联结构成一个整体(刘式达,1999;Shi,2005;Dai,2005)。国内外学者对气候系统的研究已经取得了不少进展(Xiao,2007;封国林,2005;Wang,2005),特别是近 10 多年来,随着气候研究的迅速发展,已认识到气候系统具有多层次性和非平稳性;局地气候变化依赖于全球气候变化,局地气候变化也会影响全球气候。对于气候系统内部各区域气候子系统之间作用的研究,就相互作用的空间距离而言主要有两个方面:(1)被称为遥相关作用的远程关联。其中较为典型的有 Walker 提出的北大西洋涛动(NAO)、北太平洋涛动(NPO)和南方涛动(SO)(Keppenne,1992;Lamb,1987;Walker,1983);龚道溢等(1998)发现了南极涛动;Thompson 等(1998)基于对北半球大气环流的分析,提出了北极涛动(AO),发现 AO 和 NAO 的相关系数高达 0.69。(2)大气本身是不可压缩的流体,如果将整个大气系统看成由若干个小的区域作为基本单元组成的复杂系统,这些基本单元之间存在能量、动量、热力学熵等物理量的交换,即存在一定程度的短程作用和联系(胡隐樵,2002)。

半个世纪以来,通过气象学家的探索和分析研究,尤其是 Charney(1966)、Lorenz(1969a,1969b)、Smagorinsky(1969)等的工作,人们对大气运动短期和中长期的预报问题等的认识有了显著提高。丑纪范等(2001)评述了目前短期气候数值预测方面存在的缺陷,提出需要解决的 4 个挑战性问题,指出了短期气候预测的前景和新的发展方向。洪梅等(2006)通过模型积分和 EOF 的时空重构,实现了副热带高压活动场的中长期预报。丁瑞强等(2008)基于非线性误差增长理论对 500 hPa 温度场、纬向风场以及 500 hPa 高度场的可预报性时空分布特征进行了研究。目前,从气候要素全球格点间的关联性多尺度特征的角度研究气候要素关联的短期和中长期变化问题还少有开展。因此,有必要通过合理的数理统计,并结合动力学方法对气候系统内部各子系统之间的关联性进行系统地研究,旨在对各种时空尺度的气候状态进行描述和预测。但目前这方面的研究大都局限于对少数几个区域气候子系统之间可能作用和联系进行研究,很少有工作将整个气候系统内部所有子系统之间的相互关系视作一个巨大的复杂网络系统进行研究,即从全局性的角度进行研究。由于气候系统的内部复杂性,从全局性的角度研究气候系统的内在联系可能有助于更好地理解气候系统的基本结构特征,进而挖掘其内部性质。本文采用 NCEP/NCAR 全球再分析气温资料,结合矩阵理论讨论了全球气温序列之间的关联性与随机序列之间关联性的联系和区别,研究了对格点气温序列关联性影响最显著的本征模态,初步讨论了全球气温序列之间关联性的时空演变的多尺度特征。

2 资料和方法

2.1 资 料

气温资料来源(Kistler,2001)是美国国家环境预报中心(NCEP)和美国国家大气研究中心(NCAR)对全球 1948—2005 年的气象资料进行再分析形成的格点逐日平均气温资料(地面资料),分辨率为 5°×5°,沿纬向有 72 个格点,沿经向有 36 个格点,总计 2592 个格点,每个格点对应长度为 1948—2005 年的资料为 $T_i(j)$,$j=1,2\cdots21170$;$i=1,2\cdots2592$。由 30 年温度资料计算得到各格点对应一年中每一天的温度基准序列 $\overline{T}_i(l)$,$l=1,2\cdots365$;$i=1,2\cdots2592$。为滤除日温度资料中包含的季节振荡信号,对数据做距平处理得到距平序列

$$T'_i(j)=T_i(\mathrm{mod}(j,365))-\overline{T}_i(l) \qquad (1)$$

2.2　关联性研究方法

关联矩阵理论的主要思想如下：对于一个 M 维多变量数据集合 $X_i(t)$，其中 $i=1,2\cdots M$；$t=1,2\cdots N$。首先对原始数据作标准化处理：

$$X'_i(t)=\frac{X_i(t)-\langle X_i\rangle}{\sigma_i} \tag{2}$$

其中$\langle X_i\rangle$表示平均值，$\sigma_i=\sqrt{\langle X_i^2\rangle-\langle X_i\rangle^2}$，通过标准化处理 $X'_i(t)$的均值为 0，标准偏差为 1。则关联矩阵的元素为(Murirhead，1982)

$$C_{ij}=\frac{1}{N}\sum_{t=1}^{N}X'_i(t)X'_j(t)$$
$$(i=1,2\cdots M;j=i+1,2\cdots M) \tag{3}$$

以此构建得到的关联矩阵，其每一个元素 C_{ij} 的范围为 $-1\leqslant C_{ij}\leqslant 1$，$C_{ij}=1$ 表示完全正相关，$C_{ij}=-1$ 表示完全负相关，$C_{ij}=0$ 表示不相关。计算关联矩阵的本征值 λ_i 和本征矢量 v_i，则有

$$Cv_k=\lambda_k v_k\quad(k=1,2\cdots N) \tag{4}$$

其中本征值将根据其大小进行排列：$\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_N$，且本征值的概率分布直接体现 M 维多变量数据集之间的关联程度(Müller，2005)。本文的研究中，$M=2592$，$N=21170$。

对由气候要素构建的关联矩阵中的关联系数 C_{ij} 的重要性和意义进行分析主要存在以下两个方面的局限性：(1)气候背景总是随时间变化的，任意两个格点气候要素之间的关联性并不是稳定不变的；(2)由于观测数据长度的有限性，在分析格点之间关联性时经常会伴随有一些虚假关联。怎样才能辨别各格点气候要素之间在某一时段内的关联性是真实的？为此，本文采用随机关联矩阵 $\boldsymbol{R}$ 来给定“无效关联区间”。随机关联矩阵是由若干相互之间没有任何相关性的随机序列构建得到的。如果关联矩阵 $\boldsymbol{C}$ 的统计性质和随机关联矩阵 $\boldsymbol{R}$ 一致，则说明 $\boldsymbol{C}$ 的相关性是虚假的，如果不一致则说明 $\boldsymbol{C}$ 的关联性是真实的。因此，研究中需要将关联矩阵 $\boldsymbol{C}$ 分为与随机关联矩阵统计性质一致和偏离两个部分分别进行讨论。

3　结果和分析

3.1　关联系数的概率分布

对 1948—2005 年 NCEP 格点逐日气温距平资料 $T'_i(j)$应用关联矩阵理论计算任意两格点气温序列的关联系数，进而构建气温关联矩阵 $\boldsymbol{C}$，并分析$\{C_{ij},i,j\neq 1\}$的概率分布 $P(C_{ij})$。虽然受全球变化的影响，气温序列中存在一定增暖的线性趋势，但这种趋势的存在对于计算气温序列关联矩阵的影响较小，最直接的体现就是关联矩阵中既有接近 1 的强关联，又有接近 0 的弱关联，并且弱关联在很小的区域内存在；此外，我们也尝试滤去气温序列的趋势项，然后构建关联矩阵，结果与此相似。应用随机函数产生随机序列，计算相互之间的关联系数，构建随机关联矩阵 $\boldsymbol{R}$，进而比较分析气温关联矩阵 $\boldsymbol{C}$ 和随机关联矩阵 $\boldsymbol{R}$ 之间关联性的差异。此外，考虑气候变化的尺度特征，分别对气温序列做 5、10…30 天尺度的滑动平均，进而研究各种不同尺度下格点温度的关联性。图 1a—1c 分别为随机关联矩阵、1 天和 15 天尺度下温度序列构建的关联矩阵。显然，随机关联矩阵的排列杂乱无章，毫无规律可言，关联系数非常小；1 天尺度下的关联矩阵则表现为由对角线开始关联系数向两侧递减，关联系数分布范围较广；15 天尺度下的关联矩阵和 1 天尺度的情况分布相似，但关联系数总体较大，关联一致性更好。

从 3 种矩阵关联系数概率分布(图 2a)可以看出，随机关联矩阵 $\boldsymbol{R}$ 的关联系数概率分布 $P(R_{ij})$很好地符合高斯对称分布，均值为 0。此外，其关联系数的值域为[－0.02，0.02]的狭小范围，这说明随机序列之间几乎不存在任何有意义的联系，这种分布对应的关联系数 R_{ij} 仅仅是一种没有意义的关联“噪声”。相比随机序列的关联性，格点温度序列之间关联系数的概率分布有一定的差异。就 1 天尺度而言，$P(C_{ij})$在[－0.02，0.02]的范围内也存在较大的尖峰值，这说明在 1 天等小尺度上，全球格点温度资料之间的关联存在一些虚假成分，因此在分析其相互之间关系时应该首先剔除这种关联“噪声”；但在[－1.0，－0.02]和[0.02，1.0]这一相对较宽的范围内 C_{ij} 也存在一定的分布。此外，对 1 天尺度格点气温序列随机打乱其排列顺序，然后计算各格点之间的关联系数及其概率分布(图 2a)，此时关联系数的概率分布又退化为随机序列的高斯分布。因此，格点气温资料之间必然存在一些有意义的真实关联，这种关联可能是全球气候背景下格点气温同步变化的表现。

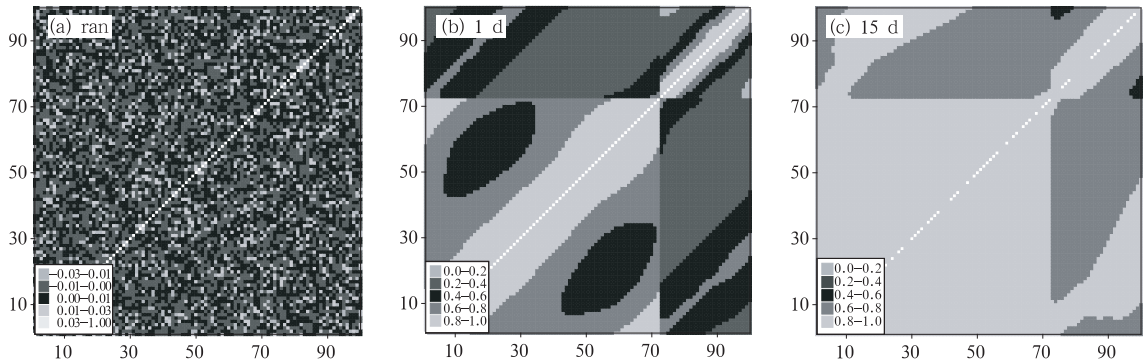

图 1 100×100 关联矩阵 ***C***

(a. 随机关联矩阵,b. 1 天尺度关联矩阵,c. 15 天尺度关联矩阵)

Fig. 1 100×100 correlation matrix

(a. the random correlation matrix, b. the 1 day scale correlation matrix, and c. the 15 days scale correlation matrix)

1—30 天尺度的关联,是既存在一定关联“噪声”,又存在一定真实关联的单峰分布,格点之间的关联相对较弱,关联系数的均值随着尺度增加而增加(图 2a),概率分布的峰值均向右略偏移。随尺度的增加(图 2b),关联系数大于 0.3 格点间关联系数的均值有所下降,但所有格点温度序列间关联系数的均值有所增加,且 15 天尺度以后的变化趋于平缓。原因可能在于,1—10 天以下尺度以近邻或次近邻等格点之间的短程关联为主,而距离较远的格点之间的长程关联则大多对应关联“噪声”等虚假信息;尺度增加至 15 天及以上尺度可能既存在短程关联,相对大的尺度能够在一定程度减弱“噪声”和外强迫等的影响,温度变化的区域同步性更强,距离较远的格点之间的长程关联比例有所增加,从而整体关联系数均值也增大。另一方面,15 天及以上尺度的平均态下,格点气温序列之间的关联主要表现为关联系数值为 0.3—0.7 的中等程度关联,因此格点间强关联系数的均值有所下降。

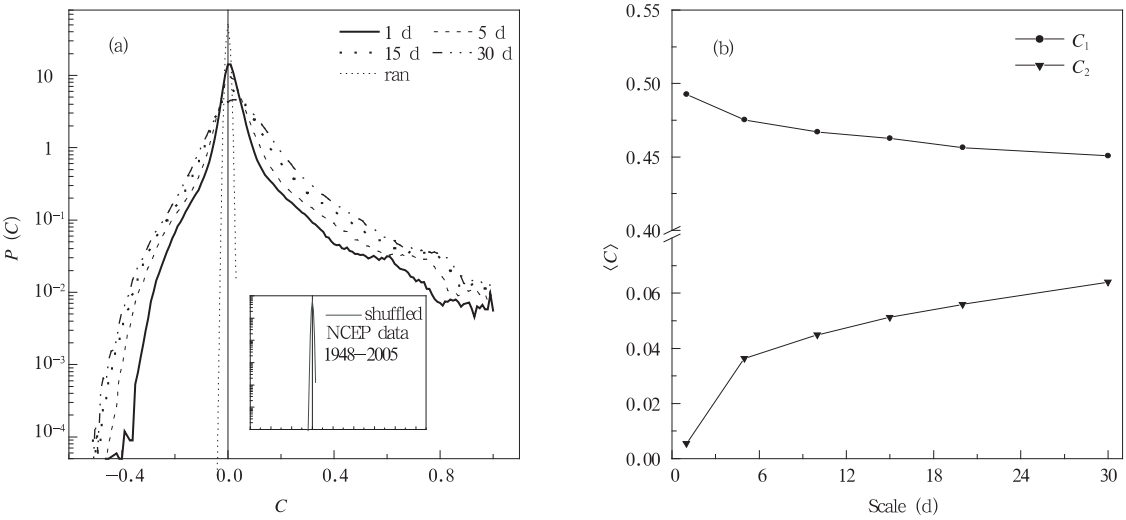

图 2 不同尺度(1—30 d)格点关联系数和随机序列(ran)关联系数的概率分布(a)以及强关联系数(C_1)和整体关联系数(C_2)的平均(b)

(图 a 中的小图为格点温度序列次序打乱后关联系数的概率分布)

Fig. 2 Probability distribution of the correlation coefficients of temperature correlation matrix on the scales 1,5,15 and 30 d as well as the random correlation matrix (a), and the means of the good correlation (C_1) as well as the all the correlation (C_2) coefficients (b)

(The small figure in Fig. 2(a) denotes the probability distribution of the correlation coefficients for shuffled temperature series)

为检验不同尺度下长程和短程关联的主次性,我们以某一格点(x_i, y_i)为中心($x_i = \left|\frac{l_i - 87.5}{5}\right|$, $y_i = \frac{z_i}{5}$),l_i 表示纬度,z_i 表示经度。其他任一格点(x_j, y_j)与其距离为 $d_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$,关联系数为 C_{ij}。我们设定关联系数的阈值 $C_0 = 0.3$,在 1 天和 15 天两种尺度情况下,分别统计了与格点(x_i, y_i)距离为 d,且关联系数 $C_{ij} > C_0$ 的关联格点数 $N_{ij}(d)$。对每个格点都进行类似处理,并将统计结果累加,可以得到与格点距离相对应的关联格点数的累计分布 $N(d)$

$$N(d) = \sum_{i=1}^{M}\sum_{j=i+1}^{M} N_{i,j}(d) \tag{5}$$

分别设定关联系数的阈值 $C_0 = 0.2, 0.4, 0.5\cdots$,多种情况的结果与 $C_0 = 0.3$ 类似。由图 3 可以看出,关联格点数累计分布与相对应的距离 d 之间满足双幂律分布。

$$P(d) = \begin{cases} Ad^{\lambda} & d < d_0 \\ Bd^{\gamma} & d > d_0 \end{cases} \tag{6}$$

其中 A, B 为常数,λ, γ 为幂律指数。1 天尺度下,$d_0 \approx 4$ 前后存在两种幂律分布,图 3 中,fit_1 表示对 1 天尺度数据点后半段的线性拟合,fit_2 则表示对前半段的线性拟合。前半段幂律指数 $\lambda = 1.13 \pm 0.07$,后半段幂律指数 $\gamma = 0.17 \pm 0.01$;15 天尺度下,$d_0 \approx 6$ 前后存在两种幂律分布,图 3 中,fit_3 表示对 15 天尺度数据点前半段的线性拟合,fit_4 则表示对后半段的线性拟合。前半段幂律指数 $\lambda = 1.27 \pm 0.074$,后半段幂律指数 $\gamma = 0.30 \pm 0.01$。由此可知,1 天和 15 天两种尺度对应了两种关联格点数累计分布类型。此外,1 天尺度,当 $d = 4$ 时,关联格点数的累计分布变化就开始趋于平缓,而 15 天尺度下则是 $d = 6$ 左右才开始趋于平缓,由此可以看出,15 天尺度下距离较远的格点之间的长程关联数较 1 天时有所增加。

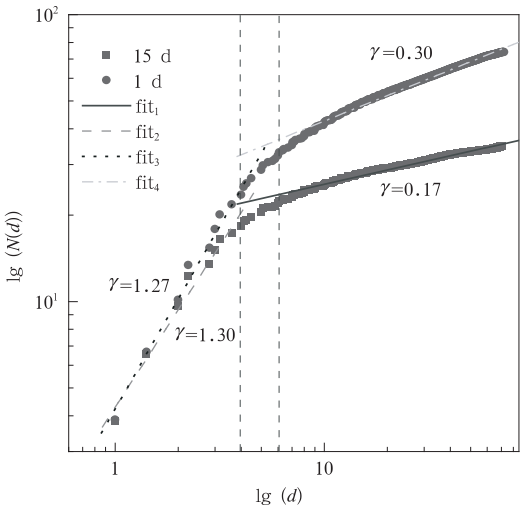

图 3　关联格点数的累计分布
(横坐标为格点间隔,纵坐标为累计关联格点数;
fit_1—fit_4 表示线性拟合(下同))

Fig. 3　Accumulative distribution of the number of correlative points
(The abscissa and ordinate are for the logarithm of the distance between two points and the accumulative correlative point number, respectively; fit_1 – fit_4 in the figure denote their linear fittings (the same below))

3.2　矩阵本征值的概率分布

上面的讨论初步表明:随机关联矩阵 $\boldsymbol{R}$ 和气温关联矩阵 $\boldsymbol{C}$ 之间存在一定差异,验证了 $\boldsymbol{C}$ 中存在关联"噪声"和真实关联等信息。但仅通过关联系数的研究显然是不够的,通过分析矩阵的本征值和本征矢量,进一步认识全球气温关联性的有关性质。

由图 4a、4b 可以看出,1 天和 15 天尺度下,气温关联矩阵的本征值概率分布 $P_{1\mathrm{d}}$ 和 $P_{15\mathrm{d}}$ 均表现为小的本征值密集出现,大的本征值稀疏出现,且本征值越大越稀疏,1 天和 15 天尺度最大本征值为 133.4193 和 1835.598。总体而言两种情况本征值的概率分布均满足指数分布特征

$$P(\lambda) = \begin{cases} 2.02\mathrm{e}^{-\lambda/0.20} + 0.04 & 1\text{ 天尺度} \\ 0.63\mathrm{e}^{-\lambda/0.18} + 0.01 & 15\text{ 天尺度} \end{cases} \tag{7}$$

出现这种分布的原因可能是关联矩阵作经验正交展开以后,一方面,本征值要满足 $\sum_{i=1}^{M}\lambda_i = M$;另一方面,一些最主要的信息可能集中在少数几个本征值较大的本征矢量中。这样就导致小的本征值大量集中出现,大的本征值稀疏出现。图 4c 为随机关联矩阵 $\boldsymbol{R}$ 的本征值概率分布 P_{ran},其分布特征显著区别于 1 天和 15 天两种情况。$\boldsymbol{R}$ 的本征值集中在一个固定的范围内。根据已有的研究(Dyson, 1971; Sengupta, 1999),完全随机关联矩阵在 $N \to \infty$, $M \to \infty$ 的条件下,$Q = N/M$ 是固定不变的,本征值概率分布满足

$$P_{\mathrm{rm}}(\lambda)=\frac{Q}{2\pi}\sqrt{\frac{(\lambda_{+}-\lambda)(\lambda-\lambda_{-})}{\lambda}} \tag{8}$$

其中 $\lambda_{-}\leqslant\lambda\leqslant\lambda_{+}$，$\lambda_{-}$ 和 λ_{+} 是随机矩阵的最小和最大本征值(Bowick，1991)。

$$\lambda_{\pm}=1+\frac{1}{Q}\pm 2\sqrt{\frac{1}{Q}} \tag{9}$$

可以得到 $Q=8.17$，$\lambda_{-}=0.42$，$\lambda_{+}=1.83$，则其本征值概率分布 $P_{\mathrm{rm}}(\lambda)$ 如图 4c。随机关联矩阵 $\boldsymbol{R}$ 的本征值概率分布特征和 $P_{\mathrm{rm}}(\lambda)$ 是基本一致的。由此可知随机关联矩阵 $\boldsymbol{R}$ 的本征值范围为 $[\lambda_{-},\lambda_{+}]$，该范围内的本征值反映了矩阵中的一些虚假关联信息。对比图 4a、4b 和 4c 发现，1 天和 15 天尺度下关联矩阵的本征值可以分为 3 个区域：Ⅰ区($-\infty$，λ_{-}]，Ⅱ区[λ_{-}，λ_{+}]，Ⅲ区[λ_{+}，$+\infty$)，Ⅰ区为小值区、Ⅱ区为随机关联区、Ⅲ区为大值区。

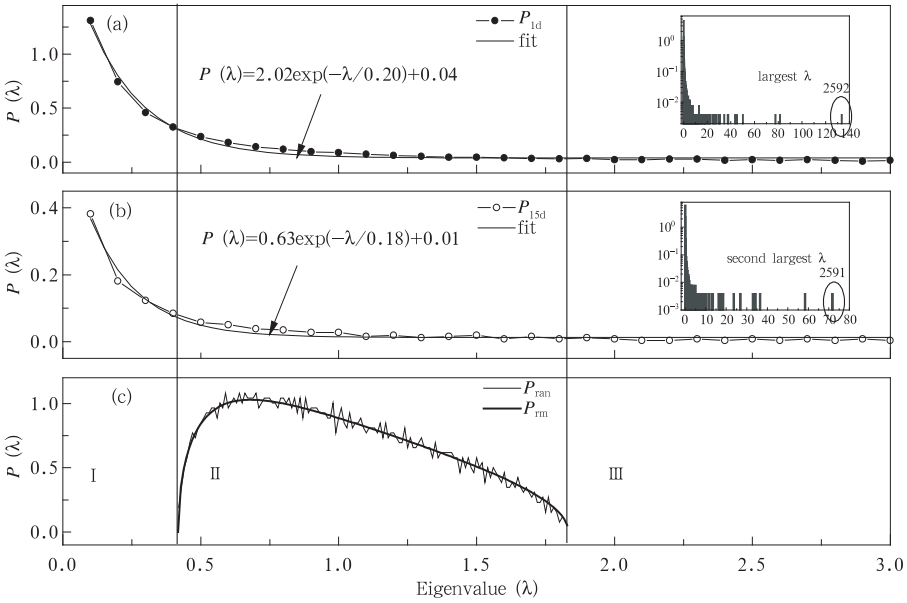

图 4 1 天(a)和 15 天(b)尺度下温度关联矩阵 $\boldsymbol{C}$ 和随机关联矩阵 $\boldsymbol{R}$(c)的本征值概率分布

Fig. 4 Eigenvalue probability distribution for the temperature correlation matrix on the 1 day (a) and 15 day (b) scales as well as the random correlation matrix (c)

3.3 矩阵本征矢量的概率分布及其意义

随机关联矩阵和气温关联矩阵之间本征值的差异，应该在对应本征矢量的特征元素中也有所体现(Laloux，1999)。通过分析本征矢量进一步解释Ⅰ和Ⅲ区内本征值可能蕴含的物理意义。由图 5 可以看出，随机关联矩阵 $\boldsymbol{R}$ 的本征值所对应的本征矢量，其特征元素的概率分布均符合高斯分布

$$P=A+B\mathrm{e}^{[-2(u-0.005)/0.04]^{2}} \tag{10}$$

其中 A、B 是可变常数。1 天和 15 天尺度下，气温关联矩阵的本征值 $\lambda<\lambda_{-}$(Ⅰ区)所对应的本征矢量 U_1 中特征元素的概率分布类似于随机关联矩阵的情况，也近似满足高斯分布。这说明Ⅰ区内的本征值及其所对应的本征矢量可能仅仅体现了关联矩阵本身的一些“噪声”，意义不大。U_{2590} 和 U_{2592} 均为两种尺度下 $\lambda>\lambda_{+}$(Ⅲ区)的本征值所对应的本征矢量。这两对本征矢量特征元素的分布均不同程度地偏离了高斯分布，并且较大的本征值所对应的本征矢量的分布偏离越明显。显然，其中必然包含了一些有别于虚假关联和“噪声”的信息。

因为较大的本征矢量可能包含格点温度序列之间的重要关联信息，因此分析Ⅰ区和Ⅲ内本征值对应的本征矢量对全球温度变化总体特征的可能影响。类似于经济序列的关联性分析(Campbell，1997；Garas，2007)，首先对于 1 天尺度而言，基于最小和最大本征值对应的本征矢量 U_1 和 U_{2592}，结合每个格点气温序列，根据式(10)构建全球气温变化总体的映射序列 $T^{1}(t)$ 和 $T^{2592}(t)$

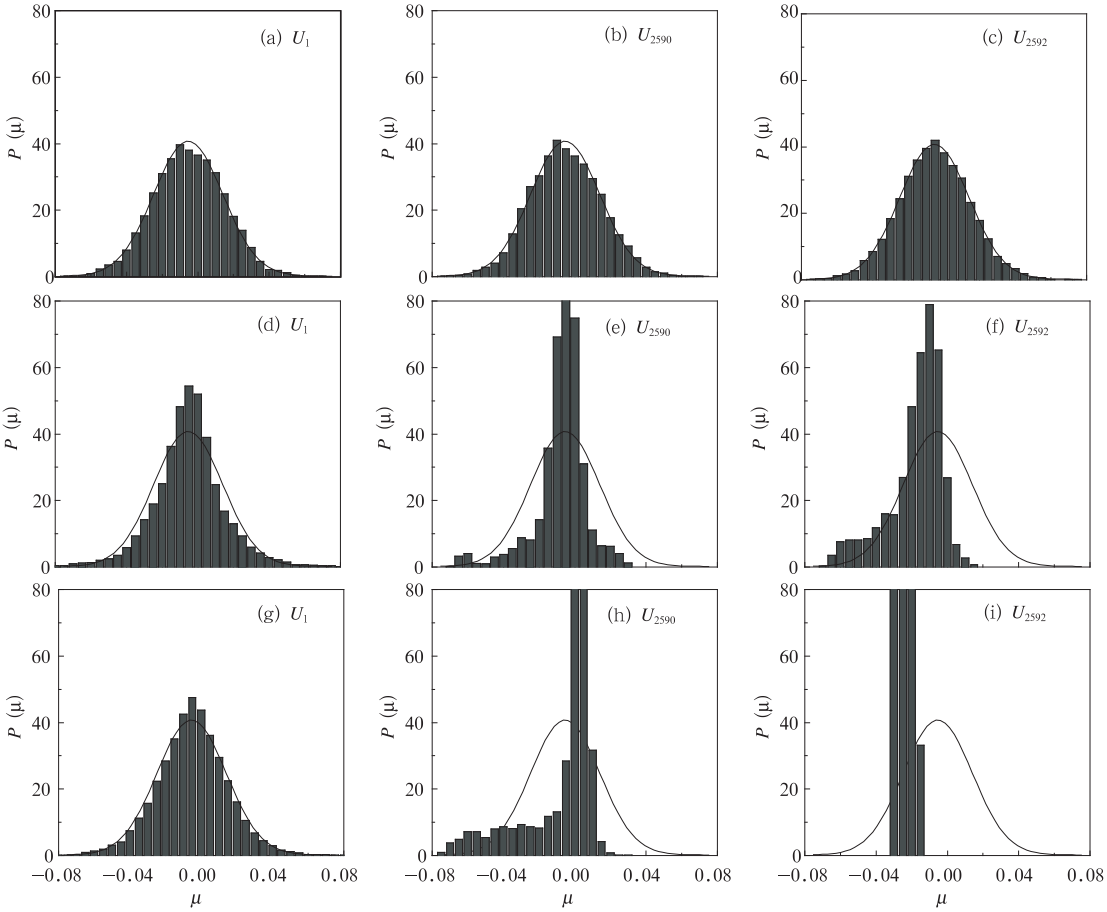

图 5　随机关联矩阵 $\boldsymbol{R}$(a—c)、1 天尺度(d—f)和 15 天尺度(g—h)下温度关联矩阵的 3 个本征矢量特征要素个数的概率分布

Fig. 5　Probability distribution of the number of the element of the eigenvectors U_1, U_{2590} and U_{2592} for the random correlation matrix (a－c) as well as the 1 day (d－f) and 15 days (g－h) scale temperature correlation matrix

$$T^i(t) = \sum_{j=1}^{M} u_j^i T_j(t) \tag{11}$$

通过这一变换,$T^1(t)$和 $T^{2592}(t)$分别体现了全球格点温度变化在本征矢量 U_1 和 U_{2592} 上的映射。此外,对所有 2592 个格点气温资料按照对应的每一天计算平均,进而得到与原序列等长度的逐日平均气温序列 $G(t)$

$$G(t) = \frac{1}{M}\sum_{i=1}^{M} T_i(t) \tag{12}$$

显然,由式(12)得到的平均气温序列 $G(t)$必然包含全球气温变化的总体特征。对 $T^1(t)$、$T^{2592}(t)$和 $G(t)$分别进行标准化处理,可以得到图 6。就 1 天尺度而言,全球气温资料在 U_1 上的映射 $T^1(t)$与全球日平均气温序列 $G(t)$的相关系数为－0.06(图 6a);而全球气温资料在 U_{2592} 上的映射 $T^{2592}(t)$与全球日平均气温序列 $G(t)$的相关系数为－0.90(图 6b)。对 15 天尺度也进行类似操作:全球气温资料在 U'_1 上的映射 $T'^1(t)$与 $G(t)$的相关系数为－0.05(图 6c);而全球温度资料在 U'_{2592} 上的映射 $T'^{2592}(t)$与 $G(t)$的相关系数为－0.99(图 6d)。由此可知,1 天和 15 天尺度下,最小本征值所对应的本征矢量包含的信息与全球的温度变化几乎没有任何关联性,最大本征值所对应的本征矢量则蕴含了全球温度变化的整体信息,$T'^{2592}(t)$可以从一个侧面体现全球温度的变化特征,U_{2592}、U'_{2592} Ⅲ区内的本征值所对应的本征矢量对格点气温之间关联的贡献较大(Plerou, 1999)。

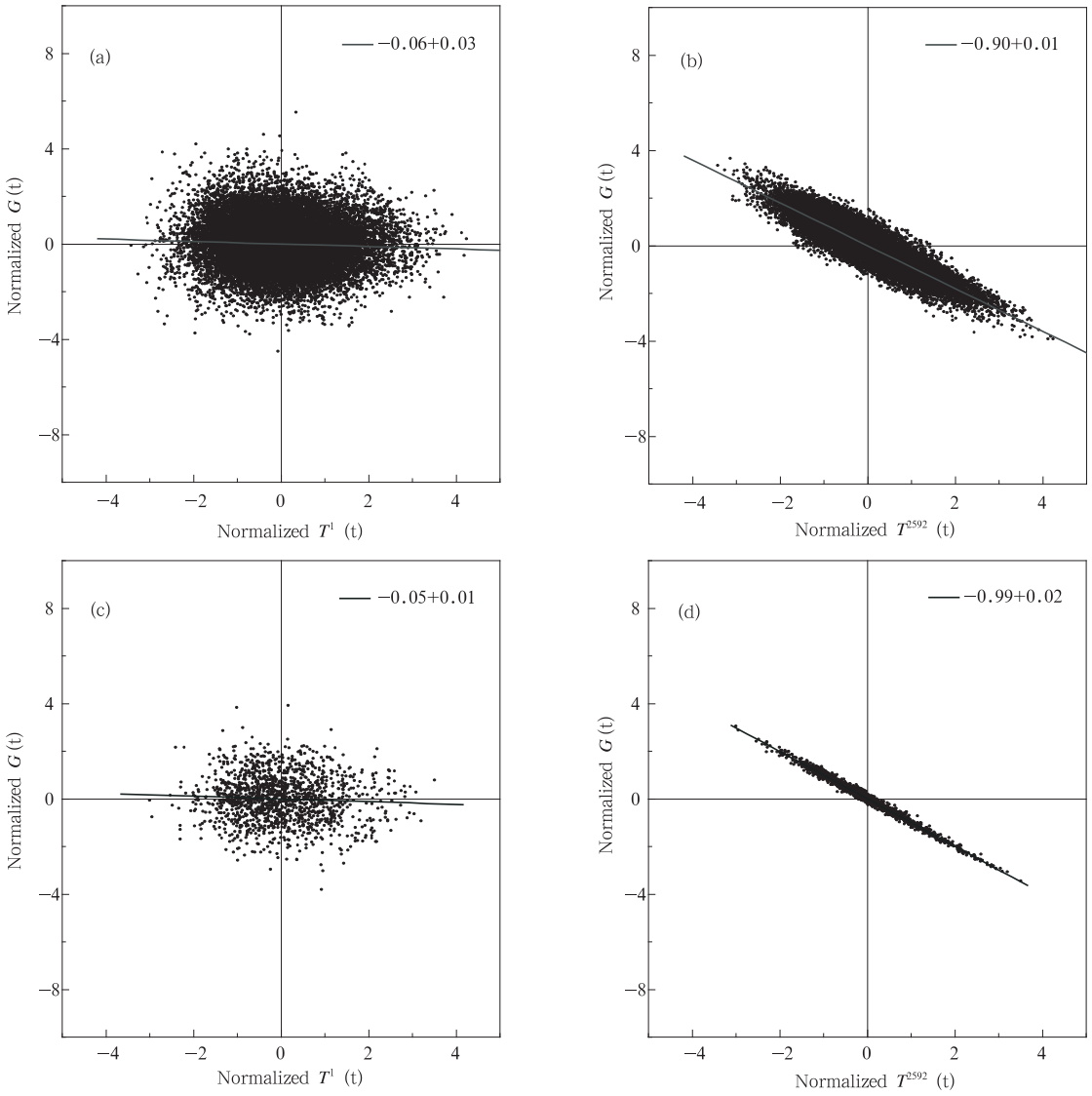

图 6 1 天(a、b)和 15 天(c、d)尺度下 $T^1(t)$、$T^{2592}(t)$和 $G(t)$的相关分布

Fig. 6 Correlations between $T^1(t)$ and $G(t)$ as well as between $T^{2592}(t)$ and $G(t)$ on the 1 day (a,b) and 15 days (c,d) scale

表 1 1 天和 15 天尺度下最大的 10 个本征值对应的本征矢量对全球格点气温关联的贡献

Table 1 The contribution to the correlation between the global temperature series of the biggest 10 eigenvectors on the 1 day and 15 days scales

	k	2583	2584	2585	2586	2587	2588	2589	2590	2591	2592	合计
1 天尺度气温关联矩阵	λ_k	30.81	34.36	37.29	37.87	43.57	45.01	50.04	77.67	81.43	133.41	—
	v_k(%)	1.19	1.33	1.44	1.46	1.68	1.74	1.93	3.00	3.14	5.14	22.40
15 天尺度气温关联矩阵	λ_k	17.40	18.47	23.78	27.00	33.07	34.27	36.47	58.44	71.71	1835.60	—
	v_k(%)	0.67	0.71	0.92	1.04	1.28	1.32	1.41	2.25	2.77	70.81	83.81
随机关联矩阵	λ_k	1.79	1.79	1.79	1.80	1.80	1.80	1.80	1.81	1.81	1.82	—
	v_k(%)	0.069	0.069	0.069	0.069	0.069	0.070	0.070	0.070	0.070	0.070	0.695

根据式(13)(Wilks, 1995),计算了 1 天和 15 天尺度下Ⅲ区较大的 10 个本征值对应的本征矢量对全球格点温度关联的贡献(表 1),其原理类似于 EOF 分解。

$$v_k = (\lambda_k \Big/ \sum_{j=1}^{M} \lambda_k) \times 100\% \tag{13}$$

显然,1 天尺度的最大本征矢量对全球格点温度关联的贡献为 5.14%,较大的 10 个本征值对应的本征矢量的贡献累计为 22.40%;15 天尺度较 1 天尺度的情况则表现的贡献更集中于少数几个较大的本征值所对应的本征矢量中:最大本征值对应的矢量贡献为 70.81%,10 个较大本征值对应的本征矢量的贡献为 83.81%。而随机关联矩阵则不存在贡献较大的本征矢量。

3.4　格点气温资料关联性的时空特征

上述研究验证了格点气温序列之间存在具有一定意义的真实关联,并且这种关联有 1—10 天尺度和 15 天及以上尺度两种类型,因此以 1 天和 15 天尺度为例进一步研究格点气温关联的时空特征。

取窗口长度 L_{win} 为 5 年,然后分别计算全球 2592 个格点在这一窗口长度内温度序列相互之间的关联系数 C'_{ij},进而构建关联矩阵 $\boldsymbol{C}'$,计算所有 C'_{ij} 的平均值$\langle C'_{ij} \rangle$:

$$C'_{ij} = \frac{1}{365L_{\mathrm{win}}} \sum_{t=1}^{365L_{\mathrm{win}}} T'_i(t) T'_j(t),$$

$$\langle C'_{ij} \rangle = \frac{1}{M(M-1)} \sum_{i=1}^{M} \sum_{j=1}^{M} C'_{ij} (i \neq j) \tag{14}$$

自 1948 年起,滑动步长为 1 年,分别计算与每个窗口对应的全球 2592 格点的关联系数平均值可以得到$\langle C'_{ij} \rangle_l$, $l = 1948, 1949 \cdots 2001$。对随机序列和次序打乱的格点气温序列也进行类似分析,图 7 可以看出,其关联系数的均值只是在 0 附近的一个很狭窄的范围内波动,基本没有特别大或特别小的区域,即完全表现出随机性,这与图 1 中两者的关联系数概率分布是一致的。对于格点气温序列而言,其关联系数的均值在1977年前后表现为两种不同变化态,这一时期的转折可能对应了 20 世纪 70 年代中后期的气温突变,1976 年的突变不仅是气温位相的变化更对应了全球格点气温序列之间关联性的变化。1948—1977 年、1987—1990 年和 1999—2002 年 3 个时段相对较高,即全球格点之间的关联性在这些时期相对较好,换而言之,全球气温

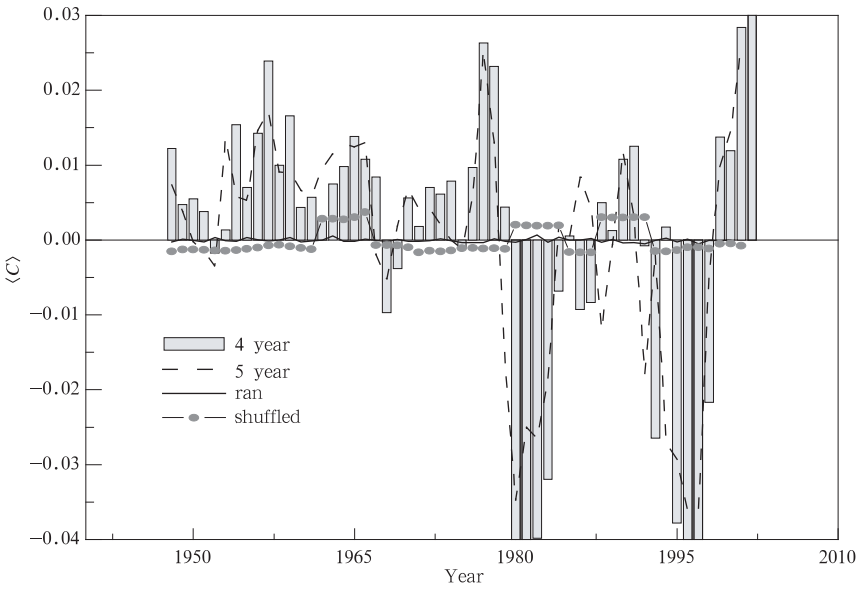

图 7　温度关联矩阵和随机关联矩阵关联系数均值距平随时间的变化

(直方图为 L_{win}取 4 年的关联系数均值,虚线为 L_{win}取 5 年关联系数,粗线为随机关联矩阵关联系数均值距平,点线为次序打乱后格点温度序列关联系数均值距平,斜线部分是关联系数相对较高的时期)

Fig. 7　Changes in the nomalies of the average correlation coefficients for the temperature and random correlation matrix with year

(Histogram, the averaged correlation coefficients for $L_{\mathrm{win}} = 4$ years, dashed line, for $L_{\mathrm{win}} = 5$ years, solid line, for the random, dotted line, for the shuffled)

变化的区域同步性相对较好；而 1978—1982 和 1991—1996 年则是均值相对较低的时期，全球气温变化的区域性差异显著。就一般而言，大部分节点之间的关联是虚假关联，这种关联“噪声”的值一般较小且相对较稳定，对于整个气候系统影响较大的主要是关键区域的一些重要的长程关联。这里关联系数较大的时段和较小的时段间隔出现，具有一定的准年代际振荡，3 个关联系数值较大的时段可能对应着某些气候变化关键区域与其他节点之间关联位相的较大变化。对关联系数的年际变化做小波分析和功率谱分析，发现其中存在 10—20 年左右的准周期振荡。根据肖栋等(2007)、封国林等(2006)的研究，气候变化存在显著的年代际特征及突变，因此在发生年代际转折和突变的过程中，气候系统内部部分节点之间的一些重要关联会发生一定的异常。为了进一步验证计算结果的可靠性，窗口长度 L_{win} 取 4 或 6 年等，结果和 L_{win} 取 5 年时的情况类似(图 7)。

计算每个格点关联系数的平均值，并给出了格点气温序列关联系数的空间分布(图 8)。1 天和 15 天尺度下，格点的关联系数均表现为沿纬向呈带状分布，关联最显著的区域为赤道及其两侧的低纬度区域。1 天尺度下，A 区对应北太平洋海域的关联异常小，原因可能在于北太平洋海域是反映全球大气和海洋变化的敏感区域，这一区域的海表温度存在太平洋年代际涛动，与南太平洋海域等可能存在显著的负相关作用(Mantua, 1997; Trenberth, 1994)，并且这些负相关作用在 1 天尺度就有一定的体现；C 区对应西太平洋暖池(李克让, 1998)、D 区对应印度洋暖池(Wang, 2001)、E 区对应热带大西洋海温异常区域(Rodwell, 1999)，根据已有的研究(殷永红，2001；Houghton，1992)，这3个区域海

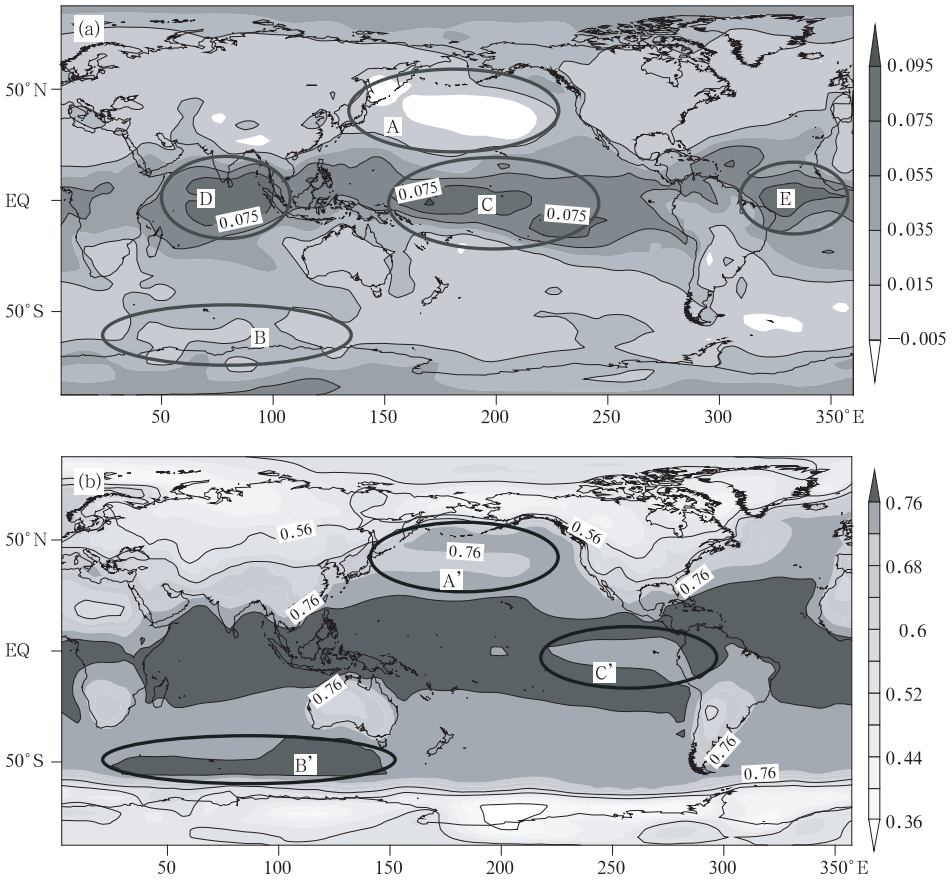

图 8 格点关联系数的空间分布(a. 1 天尺度，b. 15 天尺度)

Fig. 8 Spatial distribution of the gridpoint correlation coefficients on the 1 day scale (a) and 15 days scale (b)

表温度的变化存在一定的同步关联,显然,这种关联在1天尺度上就有一定的体现,进而加强了3个区域的关联强度。B区对应南印度洋海域,这一区域的海表温度在1天尺度上其关联性较周围区域相对偏小,即该区域的海温可能与其他某一区域的海温之间存在弱的负相关。15天尺度下,A′区对应北太平洋海域的关联仍然异常小,说明1天尺度下的负相关作用在15天尺度上依旧存在,并且负相关性有所加强;B′区对应南印度洋偶极子海域(Behera, 2001),关联较其周边海域异常大,这和1天尺度的情况正好相反,说明在不同的尺度下,区域之间的关联性可能会发生质的转变;这一区域的关联异常可以排除是短程关联强的结果,因此可能与其他区域之间存在一定的遥相关作用(Reason, 1999)。C′区(热带东太平洋海域)的关联异常小,这一海域恰好对应发生El Niño的区域,根据已有的研究(Wang, 1995; 李崇银, 1999),El Niño区域与赤道西太平洋暖池区域的海表温度存在较强的负相关作用,因而El Niño区域的关联相对较小。对于热带三大洋地区的3个暖池,15天尺度下,由于海洋系统温度变化的记忆性和均一性,短程关联的强度有所加强,从图8b中3个暖池区域的相对较大的关联强度未能很好地体现。但从图9可以看出,这3个区域的关联强度还是相对较强的,但暖池区域关联强度与周围地区相对差异减小。1天和15天尺度两种情况的差异还体现在:1天尺度下,关联最小的区域为南北半球的中纬度区域,而15天尺度下则关联最小的纬度带为包括两极在内的高纬度区域。就整体而言,海洋上的关联系数值要明显大于陆地的情况,15天尺度的情况下这一特征尤其显著。

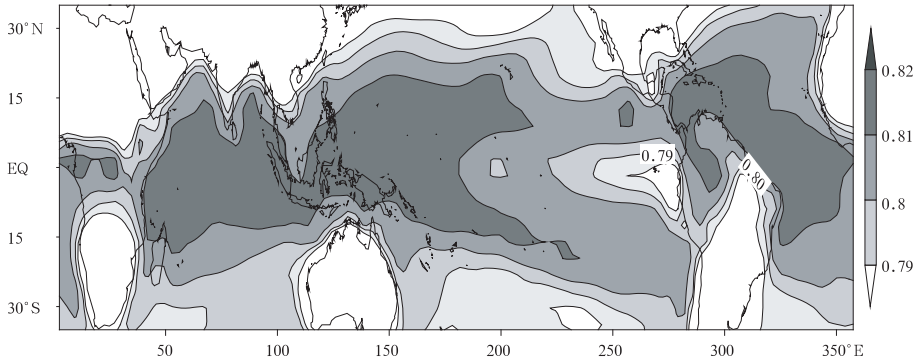

图9　15天尺度低纬度格点关联系数的空间分布(32.5°N—32.5°S)

Fig. 9　Spatial distribution of the gridpoint correlation coefficients on the 15 days scale in the low latitudes (32.5°N －32.5°S)

为进一步分析格点的纬向和经向特征,对1天和15天两种尺度的格点关联系数计算纬圈和经圈平均(图10)。1天尺度,格点关联系数纬圈平均的最大值都在赤道线上,并从赤道向两侧递减,最小值位于南北纬略小于60°区域,由60°分别向南极和北极又逐步升高,整体而言南北半球具有很好的对称性。15天尺度情况和1天尺度类似,格点关联系数的最大值也位于赤道线上,但其对称性不如1天尺度的情况好,北半球30°—60°N区域的关联系数值相对南半球大许多,关联系数的最小值位于南北半球略大于60°的纬度带上,两极区域关联性略有上升(图10a)。出现这一现象的原因可能是海洋系统比热容相对较大,下垫面差异较小,15天尺度又表现为海温变化的一种均态,因此节点之间差异较小,关联性较好。从图10b可以看出,15天尺度,格点关联系数的经向平均最大值位于10°W附近,整体上呈准对称分布,关联系数相对较小的区域位于20°—100°E和60°—120°W附近。1天尺度,格点关联系数的经向均值不再有对称分布,大体呈现波峰-波谷-波峰-波谷-波峰-波谷的特征。60°E—155°W区域,其关联系数的变化与15天尺度的位相相反,而其余区域的位相则与15天尺度的位相类似。

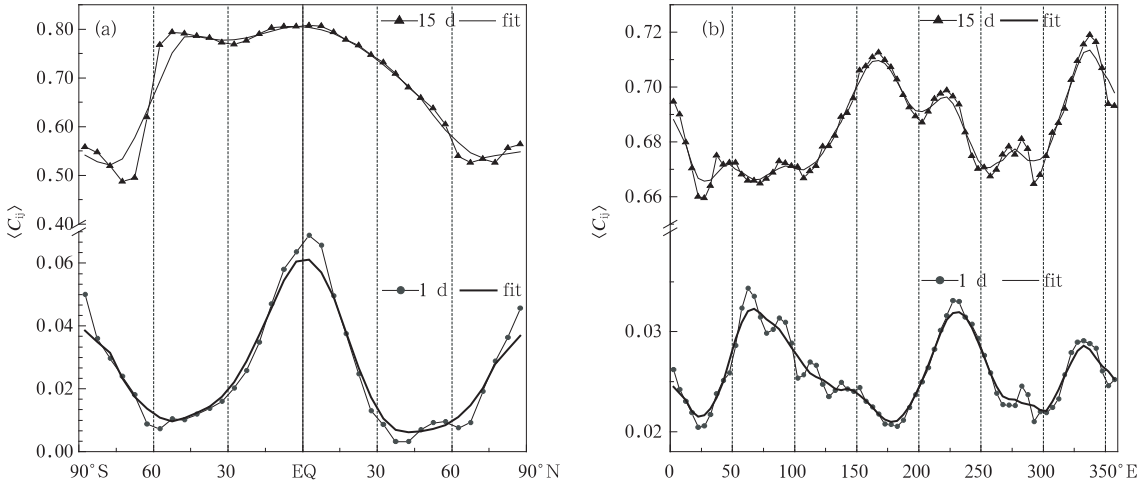

图 10 1 天和 15 天尺度格点关联系数的纬向(a)和经向(b)平均

Fig. 10 Latitudinal (a) and longitudinal (b) means of the gridpoint correlation coefficients on the 1 day and 15 days scale

4 结论和讨论

通过气温关联矩阵和随机关联矩阵关联系数、本征值和本征矢量概率分布特征的比较发现,全球格点气温序列间既存在关联“噪声”,又存在真实关联;格点气温序列之间存在的关联既包括近邻或次近邻格点之间的短程关联,又包括如太平洋年代际涛动等远程格点之间的遥相关联。对于不同的时间尺度,这两种关联的表现各有差异,1-10 天的小尺度情况下,格点之间的关联以短程关联为主导;而 15 天及以上尺度则是短程关联比例显著下降,长程关联有所上升。对关联矩阵和随机关联矩阵求解本征值和本征矢量发现,格点气温序列之间的关联信息主要包含在几个较大本征值所对应的本征矢量中,并且格点气温序列在这些本征矢量上的投影能够在一定程度上体现全球气温变化的整体特征。此外,格点气温序列关联性在时间和空间上均存在显著的变异性,主要表现为关联系数的均值在 1977 年前后表现为两种不同变化态,对应了 20 世纪 70 年代中后期的气温突变;1948—1977 年、1987—1990 年和 1999—2002 年 3 个时段相对较高,即全球格点之间的关联性在这些时期相对较好,而 1978—1982 和 1991—1996 年则是均值相对较低的时期,全球气温变化的区域性差异显著。关联系数的空间分布在两种尺度下都表现为沿纬向呈带状分布,但 1 天尺度的纬向平均具有较好的对称性,而 15 天尺度的经向平均具有一定的准对称性。因此就关联性的角度而言,1 天尺度和 15 天及以上尺度的格点气温序列可以构成两种性质不同的关联网络(汪小帆, 2005; 龚志强, 2008),而 15 天尺度下构成的复杂系统可能相对更复杂。应用复杂网络的最新研究成果进一步分析和解释这种关联网络将有助于更好地理解气候系统这一复杂巨系统的内在联系和规律,进而为我们从新的角度模拟温度、降水等气候要素的变化提供可能的途径。这些工作将在接下来的文章中陆续介绍。

参考文献

丑纪范,徐明. 2001. 短期数值预报的进展和前景. 科学通报, 46(11): 890-895

丁瑞强, 李建平. 2009. 天气可预报性的时空分布. 气象学报, 67(3): 343-354

封国林, 董文杰, 龚志强等. 2006. 观测数据非线性时空分布理论和方法. 北京:气象出版社, 230pp

封国林, 龚志强, 董文杰等. 2005. 基于启发式分割算法的气候突变检测研究. 物理学报, 54(11): 5494-5499

龚道溢, 王绍武. 1998. 南极涛动. 科学通报, 43(3): 296-301

龚志强, 周磊, 支蓉等. 2008. 1-30 d 尺度温度关联网动力学统计性质的研究. 物理学报, 57(8): 5351-5360

胡隐樵. 2002. 大气热力学导论——大气非平衡态非线性热力学. 北京: 地质出版社, 406pp

洪梅, 张韧, 何金海. 2006. 基于动力统计模型重构的副热带高压中长期预报. 气象学报, 64(6): 780-789

李克让, 周春平, 沙万英. 1998. 西太平洋暖池基本特征及其对气候的影响. 地理学报, 53(6): 511-519

李崇银，穆明权. 1999.厄尔尼诺的发生与赤道西太平洋暖池次表层海温异常. 大气科学，23(5)：513-521

刘式达，刘式适. 1999.大气动力学. 北京：北京大学出版社，536pp

汪小帆，李翔，陈关荣. 2005.复杂网络理论及其应用. 北京:清华大学出版社，260pp

肖栋，李建平. 2007.全球海表温度场中主要的年代际突变及其模态. 大气科学，31(5)：839-854

殷永红，倪允琪. 2001.热带海表温度和低层风场的年际变率及与ENSO循环的相关特征研究. 气象学报，59(4)：459-470

Bowick M J，Brézin E. 1991. Universal scaling of the tail of the density of eigenvalues in random matrix models. Phys Lett B，268：21-28

Behera S K，Yamagata T. 2001. Subtropical SST dipole event in the southern Indian Ocean. Geophys Res Lett，28：327-330

Campbell J，Lo A W，MacKinlay A C. 1997. The Econometrics of Financial Markets. Princeton：Princeton University Press，218pp

Charney J G，Fleagle V E，Lally V E，et al. 1966. The feasibility of a global observation and analysis experiment. Bull Amer Meteor Soc，47(2)：200-220

Dai X G，Fu Z B，Wang P. 2005. Interdecadal change of atmospheric stationary waves and North China drought. Chin Phys，14(4)：850-858

Dyson F J. 1971. Distribution of eigenvalues for a class of real symmetric matrices. Rev Mex Fis，20：231

Garas A，Argyrakis P. 2007，Correlation study of the athens stock exchange. Physica A，380：399-410

Houghton R W，Tourre Y. 1992. Characteristics of low frequency sea surface temperature fluctuations in the tropical Atlantic. J Climate，5(7)：765-771

Keppenne C L，Ghil M. 1992. Adaptive filtering and prediction of Southern Oscillation Index. J Geophys Res，97：20449-20454

Kistler R，Kalnay E，Collins W. 2001. The NCEP/NCAR 50 year reanalysis：Monthly means CD-ROM and documentation. Bull Amer Meteor Soc，82(2)：247-267

Laloux L，Cizeau P，Bouchaud J P，et al. 1999. Noise dressing of financial correlation matrices. Phys Rev Lett，83：1467-1470

Lamb P J，Peppler R A. 1987. North Atlantic Oscillation：Concept and application. Bull Amer Meteor Soc，68：1218-1225

Lorenz E N. 1969a. Atmospheric predictability as revealed by naturally ocurring analogues. J Atmos Sci，26：636-646

Lorenz E N. 1969b. The predictability of a flow which possesses many scales of motion. Tellus，21：289-307

Mantua N J，Hare S R，Zhang Y，et al. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. Bull Amer Meteor Soc，78：1069-1079

Murirhead R J. 1982：Aspects of Multivariate Statistical Theory. Wiley：New York，308pp

Müller M，Baier G，Galka A，et al. 2005. Detection and characterization of changes of the correlation structure in multivariate time series. Phys Rew E，71：046116

Plerou V，Gopikrishnan P，Rosenow B，et al. 1999. Universal and no universal properties of cross correlations in financial time series. Phys Rev Lett，83：1471-1474

Reason C J C，Mulenga H M. 1999. Relationship between South African rainfall and SST anomalies in the SW Indian Ocean. In J Climatol，19：1651-1673

Rodwell M J，Rowell D P，Folland C K. 1999. Oceanic forcing of the wintertime North Atlantic oscillation and European climate. Nature，398：320-323

Sengupta A M，Mitra P P. 1999. Distributions of singular values for some random matrices. Phys Rev E，60：3389-3392

Shi N. 2005. An improved south Asian summer monsoon index with Monte Carlo test. Chin Phys，14(4)：844-849

Smagorinsky J. 1969. Problems and promises of deterministic extended range forecasting. Bull Amer Meteor Soc，50：286-312

Thompson D W J，Wallance J M. 1998. The Arctic Oscillation signature in wintertime geopotential height and temperature fields. Geophys Res Lett，25：1279-1300

Trenberth K E，Hurrelle J W. 1994. Decadal atmosphere-ocean variations in the Pacific. Climate Dyn，9：303-319

Wang Bin. 1995. Interdecadal changes in El Nino onset in the last four decades. J Climate，8：267-285

Wang C，Enfield D B. 2001. The tropical Western Hemisphere warm pool. Geophys Res Lett，28：1635-1638

Wang G L，Yang P C，Lu D R. 2005. The spatio-temporal series method and tests of its predictable ability. Chinese J Atmos Sci，29(1)：178-186

Walker G T，Bliss E W. 1932. World weather V. Mem Roy Meteor Soc，4：53-84

Wilks D S. 1995. Statistical Methods in Atmospheric Sciences. London：Academic Press，375pp

Xiao D，Li J P. 2007. Spatial and temporal characteristics of the decadal abrupt changes of global atmosphere-ocean system in 1970s. J Geophys Res，112 (D24S22)：doi：10.1029/2007JD008956