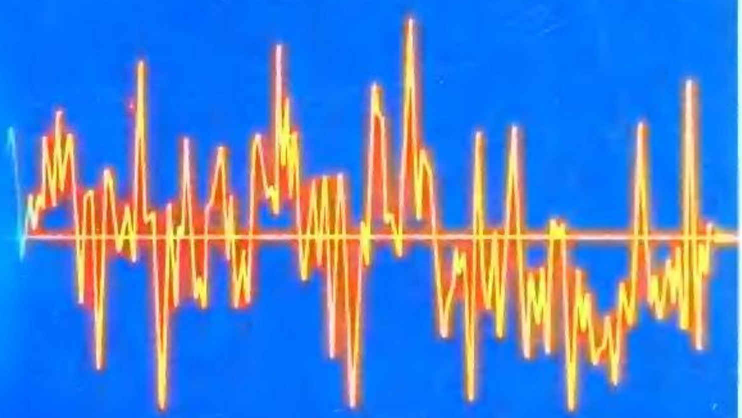
气象数据时间序列信号处理

丁裕国 江志红 编著



气象出版社

气象数据时间序列

信号处理

丁裕国
江志红 编著

气象出版社

内容简介

本书简要阐述气象数据时间序列的信号处理方法和理论。

全书共分八章：第一章简要叙述时间序列分析的理论基础；第二章详述气象时间序列的预处理方法；第三、四章从时域观点阐明描述时间序列的各种线性或非线性模型及其拟合方法；第五章则从频域观点阐明各种谱分析技术及其气象应用；第六章概略介绍多维时间序列的时频域分析，并引进滤波和线性系统的概念；第七章专门介绍时间序列模型基础上的预报方法；第八章引进动态系统分析技术，介绍几种自适应时变模型的建模方法及小波分析的气象应用。

本书可供气象科研、业务人员和有关院校师生阅读，尤其适合于已有数理统计或气象统计基础知识的初学者和本科大学生阅读；亦可供农林、水利、地震、地质、建筑、环保、交通、通讯等相关学科部门的科技人员参考。

图书在版编目(CIP)数据

气象数据时间序列信号处理/丁裕国等编著. —北京：气象出版社，1998. 1

ISBN 7-5029-2336-5

Ⅰ.气… Ⅱ.丁… Ⅲ.气象数据—时间序列分析 Ⅳ.P416

中国版本图书馆 CIP 数据核字(97)第25540号

气象数据时间序列信号处理

丁裕国　江志红　编著

责任编辑：史秀菊、顾仁俭　终审：周诗健

封面设计：姚发军　责任技编：吴　林　责任校对：史　正

气象出版社 出版

(北京海淀区白石桥路46号　邮编：100081)

北京昌平环球印刷厂印刷

新华书店总店北京发行所发行　全国各地新华书店经销

开本：850×1168　1/32　印张：9.25　字数：240千字

1998年1月第一版　　1998年1月第一次印刷

印数1～2000　　定价：17.00元

ISBN　7-5029-2336-5/P·0859

序

本人从事气候学的教学与科研已历六十余载，一向努力以统计学理论，促使统计气候学发展为气候学的主要理论基础。作者编著出版的这本《气象数据时间序列信号处理》正属于统计气候学的范畴。本人为统计气候学的发展而高兴，并衷心祝贺此书的出版。借此机会，对统计气候学在气候数值化过程中所起的重要作用，略加阐明。

气候学和天气学的数值化理论与方法有根本性的差异。短期的天气变化可主要依据确定性动力方程对高层天气变化进行数值模拟，而气候变化的数值模拟主要依据随机动力模式为宜，因为气候变量是随机变量。在气候应用方面，地面气候变化规律尤为重要，而且长期（逐年、逐月、甚或逐日）气候时空变化特征，分型分类以及分区气候预报，也正是统计气候学研究的内容，所以，气候数值化研究的任务应以统计气候学为主，努力实现统计模拟和动力模拟相结合的统计动力数值模拟。

长期气象资料记录序列实质上是各种各样随机变化着的逐次短期天气型在长期历程中、在大范围地区内随机转换的样本形象。它们正是本书作者所提气象数据时间序列信号处理方法 的处理对象，其结果必然是天气和气候预报所极需的。不过，这样的样本形象记录太庞大了，在一般气候学问题中只能把地面气象记录叠加后又加以平均。对这种复杂重叠、平均后又有随机扭曲的时空样本形象，只有以气候统计理论与统计模拟为主，去抽取分析原有的样本形象才最为适宜。气候统计模拟不仅能获得时空变化原有规律，还能把统计模拟结果作为气候形成的信息，去推论气候形成的过程。

本书作者不仅在气象数据时间序列处理的应用上作了系统的

阐述，而且在气候统计学理论上作出必要的创造和贡献，尤其能为天气和气候预报的决策提供方法，这是作者多年来在这一领域教研有素、竭力奉献的结果。

么枕生
1997年9月5日于
南京大学

前　言

气象工作者是以大气环境作为其实验场所，而将大气系统及其运动和变化作为其实验对象的。因此，气象观测数据记录是极其珍贵的实验数据记录。它们往往是一连串按时间推移而变化的数据记录，这种依时间坐标来排列的记录序列，通常称为数据时间序列。如同其它自然科学数据一样，气象数据时间序列蕴含着极其丰富的信息，它们本质上反映了大气运动和变化的物理学与动力学规律。

现代理论已经证明，实际地球大气-气候系统是一个远离平衡态的复杂系统，任何气象时间序列所包含的丰富信息都可能是参与其中的全部变量的影响结果。人们必须根据各自的研究目的和需要，设法提取那些反映规律性的气象信号，而排除混杂其中的“噪声”干扰，这种对数据时间序列提取有用信号的处理过程就是时间序列分析。一般它包括两方面内容：(1)时域上的模式模拟与预报，简称时域分析；(2)频域上的各种谱分析，简称频域分析。两者虽有异曲同工之效果，互为联系，但又各有其自身特点，并不能相互替代。

事实上，时域、频域分析两者都是当今统计科学及其应用领域发展最快的分支。谱分析技术用于气象科学的历史可追溯到本世纪40年代，早期的气象和气候研究中就已引入最简单的谐波分析作为工具，其后由于大气湍流研究的需要而广泛采用经典的湍谱分析技术，从而促进气象学科各分支对于谱分析技术的高度重视。70年代以来，现代谱分析技术和自动控制理论的发展更加拓宽了谱分析的理论和应用，与此同时，各种时间序列概率统计模式模拟和预测理论的迅速发展又为气象科学各分支，尤其是各种时间尺度的天气气候预报问题提供了重要的研究工具。现在，无论在天气

动力、气候、农业气象、大气环境或大气探测等各个学科都可以看到应用时间序列和谱分析技术所发挥的作用。

本书是作者自1985年以来为在南京气象学院开设“气象应用谱分析与时间序列”课程所编讲义基础上，经10多年教学实践逐年修改补充而成。为便于讲授与自学相结合，原讲义力求避免抽象的纯数学体系和一般文献综述，注重学科的系统性、概念的准确性和深入浅出的叙述方式，强调方法的实用性和可操作性。本次修改出版，除增补近几年国内外最新成果，引进一些新技术方法外，还针对气象学科特点，补充了许多气象应用实例，有一些是作者多年科研实践的成果。

全书共分八章。其中第一、三、四章及第二、五章的大部分由江志红执笔，第二、五章部分和第六、七、八章由丁裕国执笔，初稿写成后由丁裕国统编润笔。书中大部分实例计算和绘图由江志红完成。

本书编写过程中，曾得到南京气象学院各级领导的大力支持，屠其璞教授和施能教授为本书提出宝贵意见，阮均石教授为本书出版积极提供方便，气象出版社顾仁俭先生、陶国庆先生对本书的出版给予大力协助和支持，在此一并致谢。

作者尤为感谢我国气象学界老前辈、著名气候学家么枕生教授欣然为本书的出版，作序祝贺并热情鼓励。

鉴于学科发展迅速，作者学识水平所限，错漏和疏忽之处在所难免，敬请读者指正。

作　者

1996年12月于南京

目　录

第一章　数据时间序列分析的理论基础

几乎一切自然的和社会的科学实验数据都必须经过适当的加工整理，从中提取各自所需要的有用信息，才能透过现象，分析本质规律，预测事物的可能发展，并揭示其形成原因。气象观测数据是一种极其珍贵的实验数据，它们往往是一连串依时间推移而变化的数据记录，这种依时间坐标而排列数据的记录序列，通常称为数据的时间序列。气象数据时间序列如同其它自然科学的数据时间序列一样，蕴含着极其丰富的气象信息，人们必须根据各自的研究目的和需要，设法提取那些具有规律性的气象信号而排除混杂其中的“噪声”干扰。这种对数据时间序列提取有用信号的处理过程就是时间序列分析。本章首先简要地阐明数据时间序列分析的理论基础和重要的基本概念。

§1.1　随机过程与时间序列

大多数读者都已具备概率论与数理统计知识，众所周知，“概率统计”研究的对象为“随机变量”。假如对同一随机变量的变化过程独立地重复进行多次观测或对该随机变量变化的全过程进行一次观测，将会有怎样的结果呢？显然，前者必然获得多个以时间 t 为自变量的（非随机）函数，后者必然获得一个以时间 t 为自变量的（非随机）函数。倘若以时间 t 为横坐标，以随机变量的取值为纵坐标，必可绘制出一簇时间函数曲线。以气象变量的观测为例，若对某地历年 5 月份逐日平均气温进行上述观测（假定这些年份该地气候条件不变），就可得到一簇（5 月份）逐日平均气温的变化曲线。显然，这就意味着，对某一固定日期，气温的取值为随机变量，

而对于某一年的 5 月份，逐日平均气温的取值又构成时间 t 的函数。上例仅作为一个直观的引例，实际观测结果当然并不如此简单。综上所述，随机变量若同时随着另一参数(如时间 t) 而演变，则必然组成多个随机变量，假定每一时刻都对应着一个随机变量，必然组成无穷多个随机变量，这种随着另一个参数(如时间 t)而变化着的随机变量就称为随机过程，又叫做随机函数。

一般说来，“概率统计”仅仅涉及到个别随机变量或有限个随机变量(即随机向量)，而“随机过程”乃是兼有随机变量与普通函数双重属性的一簇无穷多个随机变量的集合，因而其概念和性质就更为复杂。现在，我们给出它的严格定义。

一、随机过程

定 义：一簇按参数 $t \in T$ 变化的随机变量，称为随机过程，记为 $\{X(t), t \in T\}$，简记为 $\{X(t)\}$。通常 T 总是取下列集合之一：

(1) $T = (-\infty, +\infty)$ 或 $T = [0, \infty)$

(2) $T = \{\cdots, -2, -1, 0, 1, 2, \cdots\}$ 或 $T = \{1, 2, \cdots\}$

若 T 取(1) 的情形，则称 $\{X(t)\}$ 为连续参数过程；若 T 取(2) 的情形，则称 $\{X(t)\}$ 为离散参数过程或随机序列，这里，T 又称为参数集，一般都指时间 t (当然亦可指其它参数如空间坐标)。显然，连续参数过程 $\{X(t)\}$ 的观测结果是 $t \in T$ 的连续域上的数值记录(时间上的连续函数)，而离散参数过程 $\{X(t)\}$ 的观测结果则是一个有序数值集合。通常为区别两者，将离散参数过程简记为 $\{X_t\}$，而将连续参数过程仍记为 $\{X(t)\}$ 。本书如无特别说明，皆以此为准。

上述定义表明，任一随机过程 $\{X(t)\}$ 必然兼有随机变量与普通(非随机) 函数的双重性质。即一方面，对于某固定时刻 t_0 来说，过程 $\{X(t)\}$ 表现为随机变量 $X(t_0)$，这一固定时刻 $t = t_0$ 所对应的随机变量就称为该随机过程的一个“截口”；另一方面，对某一次观测(试验)而言，过程 $\{X(t)\}$ 又表现为时间 t 的(非随机) 函数，即

由各个时刻的一种可能取值所构成的样本函数，若记为 $x_i(t)$，$i = 1,2,\cdots,N$，则全部样本函数 $i = 1,2,\cdots,N$ 的集合必然构成随机过程 $\{\boldsymbol{X}(t)\}$ 的全体。换言之，无论从随机变量的观点或随机试验的观点来说，$\{\boldsymbol{X}(t)\}$ 的所有可能取值的集合就称为这个过程的全体或总体(*ensemble*)，而过程 $\{\boldsymbol{X}(t)\}$ 的每次观测记录(样本函数)通常称其为过程的现实(*realigation*)，可见，随机过程总体就是各个现实之集合。因此，无论从随机变量集合或现实集合来理解随机过程都是等价的。

二、时间序列

在实际工作中，常常不需要了解或不可能得到连续参数过程的全部可能记录，而只能获得离散参数过程的观测记录或连续参数过程的离散化采样记录。例如，气象记录中，采用自动化或自记仪器在时间坐标上连续读取或记录的气象要素变化曲线，虽然在连续时间域内都有记录，但用于分析时通常总是在离散时间间隔上加以采样，使其成为离散化数字记录；另一方面，大量的气象观测记录，通常总是在连续时间域内按各种固定时制所规定的时刻或各个等距时间间隔上记录气象变量的观测值，这就是通常的定时记录序列或经统计整理后的记录序列(如日、月、年平均值或总量)。所有这些时间上有序的气象记录，都可认为是气象随机过程的观测结果，即样本函数。它们的共同特点是，以离散时间为参数的数据记录序列或数字记录序列。通常把这种记录序列叫做时间序列，有时又称随机序列(它不一定以时间 t 为参数)。对照前已述及的“随机过程”概念，不难看出，时间序列就是一种离散参数随机过程或者是它的观测记录。因此，一般地说，时间序列是指在离散参数(时刻) $t_1 < t_2 < \cdots < t_i < \cdots < t_n$ 上得到的以 t 为自变量(离散参数)的有序数值的集合，并记为

$$X(t_1), X(t_2), \cdots, X(t_i), \cdots, X(t_n)$$

有时简记为 $\{\boldsymbol{X}_t\}$ 或 $x_1, x_2, \cdots, x_t, \cdots, x_n$ 。由于在实际工作中离散

参数随机序列应用十分广泛，对它们所使用的许多行之有效的统计方法，其理论基础虽来源于随机过程，但在数学处理上却常常可避免某些复杂性。因而在自然科学和社会科学的许多领域，实用上都是以"时间序列"作为其研究对象。

值得指出的是，时间序列与其观测结果是两个不同的概念，前者乃指离散参数随机序列即它的所有现实的集合；后者则是离散参数随机序列的一次现实或一个样本序列。在本书各章节的学习中，读者必须联系上下文，分辨其具体含义。

以上所述时间序列仅仅是单一变量的时间序列，即每一时刻对应一个随机变量。如果考察对象不止一个变量，而是一组多个变量，其变化过程的数据记录序列就成为一个向量时间序列。例如，按时间顺序同时记录气温、气压、湿度三种要素取值，就构成一个三维时间序列，每一时刻对应着一个三维向量(随机向量)。关于这种多变量随机过程或多维时间序列，本书第六章将专门介绍。广义地说，数字时间序列还应包含那些前后记录取值完全可以相互确定的序列，即所谓确定性时间序列。这类确定性序列实质上就是时间坐标的确定性函数，例如 $x(t) = A\cos(\omega t + \varphi)$，其中 A,ω,φ 均为常数，只要知道任意三个时刻的 $x(t)$ 值，便可定出 A,ω,φ 三参数，整个序列没有任何不确定性。为了理论研究的方便，有时需要引入这类确定性时间序列。不过，气象科学中涉及的时间序列主要是随机时间序列。今后若不特别声明，本书中的序列都是指后者而言。

§1.2 随机过程的一般性质

一、随机过程的概率分布与统计特征

一个随机变量或随机向量的统计特性完全由它的分布函数或分布密度确定，同样，一个随机过程的统计特性也完全由它的概率

分布所决定。由于随机过程是由无穷多个随机变量构成，它的全部统计特性则需由任意有穷多个时刻 $t_1,t_2,\cdots,t_m$ 所相应的随机变量 $X(t_1),X(t_2),\cdots,X(t_m)$ 之联合分布函数 $F(x_1,x_2,\cdots,x_m;t_1,t_2,\cdots,t_m)$ 来确定。

$$F(x_1,x_2,\cdots,x_m;t_1,t_2,\cdots,t_m) = P\{X(t_1) < x_1, X(t_2) < x_2,\cdots,X(t_m) < x_m\} \tag{1.2.1}$$

通常称 $\{F(x_1,x_2,\cdots,x_m;t_1,t_2,\cdots,t_m)$；任意正整数 $m\}$ 为相应随机过程 $\{\boldsymbol{X}(t)\}$ 的有穷维分布族。

虽然有穷维分布族完整地描述了随机过程的全部统计特性，但是随机过程的有穷维分布族一般难以完全掌握，即使知道了也不便于分析和实际应用。因此，实用上，不需要描述其全部统计特性，而只需掌握其主要统计特征即可。这正如在随机变量的统计描述中往往并不需确知其概率分布函数而仅知道其主要统计特征（如均值和方差）就可满足一般应用需要一样，为了基本描述随机过程的统计特性，往往只要掌握它的最主要统计特征量就可满足应用的需要。这些统计特征量主要包括均值函数、方差函数和自协方差函数等。

1. 均值函数

一个随机过程的均值（即数学期望）也是时间 t 的函数，其定义类似于一维随机变量，即有

$$\mu_X(t) = E[X(t)] = \int_{-\infty}^{+\infty} x\mathrm{d}F(x,t) = \int_{-\infty}^{+\infty} xf(x,t)\mathrm{d}x \tag{1.2.2}$$

上式表明，随机过程的均值函数就是过程在每一时刻的随机变量 $X(t)$ 的均值，它表示各时刻变量取值的集中位置。每一样本函数（或称现实）都围绕着它摆动。

2. 方差函数

方差函数反映了随机过程每一截口取值的变动情况，即相对

于均值函数的离散程度。如同随机变量一样，我们有

$$D_X(t) = \mathrm{var}[X(t)] = E[X(t) - \mu_X(t)]^2$$
$$= \sigma_X^2(t) \tag{1.2.3}$$

3. 自协方差函数

为了分析随机过程 $\{X(t)\}$ 中不同时刻随机变量间的统计联系，需要对任意两个时刻 t 和 s 考虑 $X(t)$ 与 $X(s)$ 的相关关系。类似于随机变量之间的协方差，随机过程 $\{X(t)\}$ 的自协方差函数定义为

$$R_X(t,s) = \mathrm{cov}[X(t), X(s)]$$
$$= E\{[X(t) - \mu_X(t)][X(s) - \mu_X(s)]\} \tag{1.2.4}$$

显然，$R_X(t,s)$ 为 t,s 的非随机函数。

与一般随机变量相似，为了消除变量本身量级的影响，采用它的标准化形式更为适当，因此可定义

$$\rho_X(t,s) = \frac{R_X(t,s)}{\sigma_X(t)\sigma_X(s)} \tag{1.2.5}$$

为随机过程 $\{X(t)\}$ 的自相关函数(或标准化自协方差函数)。

很明显，自协方差函数有下列性质：

$$\begin{cases} R_X(t,s) = R_X(s,t) \\ \rho_X(t,s) = \rho_X(s,t) \end{cases} \tag{1.2.6}$$

特别当 $t = s$ 时，自协方差函数

$$R_X(t,s) = R_X(t,t) = E[X(t) - \mu_X(t)]^2$$
$$= \sigma_X^2(t) \tag{1.2.7}$$

可见，$\{X(t)\}$ 的方差函数是自协方差函数的特例。因此，随机过程的主要统计特征可概括为均值函数和自协方差函数。

二、随机过程的运算

随机过程兼有随机变量和普通函数的双重性质。在时间轴上，它就是一簇随机变量序列。由于它的随机性，其运算不同于一般变量或函数。因此有必要在新的意义下，理解有关它的极限、连续性、

导数及微积分等概念。

1. 均方极限与均方连续性

回顾普通变量序列极限的概念，若对于任意小的 $\varepsilon > 0$，都有正整数 $N = N(\varepsilon)$ 存在，使得对于一切的 $n > N$，有不等式

$$|x_n - a| < \varepsilon \tag{1.2.8}$$

成立，则称序列 $x_1, x_2, \cdots, x_n, \cdots$（记为 $\{x_n\}$）以 a 为极限，即序列 $\{x_n\}$ 收敛于 a，并记为

$$\lim_{n\to\infty} x_n = a \tag{1.2.9}$$

但是，对于随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$（或简记为 $\{X_n\}$）我们不能用上述方式定义其极限，必须给出新的意义下极限的概念。根据概率论的大数定律，随机变量的收敛概念，通常有“依概率收敛”、“以概率 1 收敛”和“均方收敛”三种方式，这是随机变量不同于普通变量的极限概念。其中第三种收敛“均方收敛”是最有代表性的。我们以此来说明随机变量序列极限的概念。

若对于任意小的 $\varepsilon > 0$，当 $n \to \infty$ 时，随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 有不等式

$$E[|X_n - X|^2] < \varepsilon \tag{1.2.10}$$

成立，则称随机变量序列 $\{X_n\}$ 均方收敛于 X，或称 X 为随机变量序列 $\{X_n\}$ 的均方收敛极限。并记为

$$\lim_{n\to\infty} E[|X_n - X|^2] = 0 \tag{1.2.11}$$

或

$$\underset{n\to\infty}{\mathrm{l \cdot i \cdot m}}\, X_n = X \tag{1.2.12}$$

为了引入随机微积分的概念，必须定义随机过程的连续性（它也不同于普通函数的连续性）。

如果对于任意小的 ε，都有 $\delta > 0$，当

$$|t' - t| < \delta \tag{1.2.13}$$

时，有不等式

$$E[|X(t') - X(t)|^2] < \varepsilon \tag{1.2.14}$$

成立，则称随机过程 $\{X(t)\}$ 在 t 点连续。对某一区域 T，$t \in T$，若