中长期水文预报

范钟秀　编著

河海大学出版社

目　　录

第一章 绪 论

第一节 中长期水文预报的概念

根据前期水文气象要素，用成因分析与数理统计的方法，对未来较长时间的水文要素进行科学的预测，称为中长期水文预报。视预报的对象与内容的不同，通常把预见期在 3 天至 15 天的称为中期预报；15 天以上一年以内的称为长期预报；一年以上的则称为超长期预报。目前开展的中长期水文预报包括径流、江河湖海的水位、旱涝趋势、冰情及泥沙等预报项目。对径流预报而言，预见期超过流域最大汇流时间的即为中长期预报。

早在我国汉代，就已有关于水文情势长期变化的记载。《史记・货殖列传》有"故岁在金，穰；水，毁；木，饥；火，旱，……六岁穰，六岁旱，十二岁一大饥。"之语。反映了当时黄河流域一带旱涝年景交替出现的变化趋势。至公元 11 世纪的宋代，《宋史・河渠志》更载有："自立春之后，东风解冻，河边人候水，初至凡一寸，则夏秋大汛当至一尺，颇为信验，故谓之"信水"。说明当时已经出现了由前期水文情况估计后期水文趋势的思想，标志着长期水文预报进入了萌芽时期。至于近代的中长期水文预报是 19 世纪末、20 世纪初才开始出现的。"世界天气法"首先应用于尼罗河下游的春汛洪水预报，其后又较广泛地应用到欧洲与北美一些国家。30 年代，中国气象学家涂长望根据前期东亚大气活动中心的特征作了预测后期长江水文状况的研究。50 年代，内蒙古自治区水文总站根据杨鉴初 1951 年提出的"历史演变法"作了黄河的长期洪水预报；原水电部水文局分析了高空气象因素对后期水文情势的影响，研究了华北地区中小河流的中期预报；长江水利科学研究院根据东亚大气环流的前后期演变规律作了长江的长期水文预报。这些方法在中国的其他流域或地区也得到应用。60 年代以后，由于气象学、海洋学等其他相关学科的发展以及新的探测手段的出现，进一步发现了一些影响江河水量变化的因素。概率统计的发展则为揭示要素自身演变规律、辨别各种影响因素的显著性以及影响因素之间的相互关系方面提供了有力的工具。这些，均使中长期水文预报获得了进一步发展。

第二节 中长期水文预报的现状

随着社会主义现代化建设事业的不断发展，国民经济各个部门对水文预报提出的要求越来越高，不仅要求有正确的短期预报，而且要求有预见期更长的中长期预报；不仅要求定性预报，而且要求定量预报。实际上，从防洪抗旱的指挥，大中小型水利、水电、水运工程的兴建、管理运行直至国防建设都要求水文部门能提供预见期长、准确性高的中期与长期预报。为了满足这些要求，全国自 1958 年以来已有十多个流域机构、省市自治区水文总站以及为数更多的地区水文分站、水库管理单位以至基层水文测站都逐步开展了这一工作，并在防洪抗旱斗争中取得了一定的成绩。但是，由于影响因素的复杂与目前科学水平的限制，中长期水文预报还处于探索、发展阶段。预报精度还不能满足各个生产部门的需要。一般来讲，大

面积旱涝趋势的定性预报有一定的参考价值，而定量预报的误差还较大，特别对特大的洪涝、干旱还缺乏有效的预报能力。

中长期水文预报是介于水文学、气象学与其他学科之间的一门边缘学科。因此，要提高预报精度就必须开展多学科的协作，进一步弄清影响长期水文过程各种因素的物理本质以及它们之间的内在联系。特别要着重研究引起大旱大涝的环流异常原因及其演变规律，因为这是提高中长期预报精度的关键。事物总是可以认识的，水文要素中长期变化的规律也是可以逐步掌握的，只要发扬“科学无险阻，苦战能过关”的精神，共同努力，迎难而上，不断总结，不断提高，就一定能够把中长期水文预报的水平不断推向前进。

第三节　现行中长期水文预报的主要方法

由于预见期的增长，中长期水文预报在方法上显然已无法利用实测降水资料通过产流汇流计算或应用上下游关系来获得预报结果，必须考虑影响水文过程的各种因素或分析水文要素自身演变的规律来进行预报。目前常用的方法主要有以下几种。

一、长期预报方法

(1)用前期环流进行预报。用前期环流进行预报的方法称为天气学方法。它的基本出发点是大气环流是制约着一个流域或地区上水文情况的主要因素。大范围的洪涝与干旱总是与大范围的环流异常联系在一起的。因此，只要分析寻找出异常环流的演变承替规律，就能利用这种规律对后期水文情况进行预报。这种方法是对大量的历史气候资料，主要是高空环流的逐月平均形势与对应的水文要素进行分析综合，概括出旱涝年前期的环流特征模式，然后由前期环流特征作出后期水文情况的定性预报。或在前期月平均环流形势图上分析与预报对象关系密切的地区与时段，从中挑选出物理意义明确、统计贡献显著的预报因子，然后用逐步回归或其他多元分析方法与预报对象建立方程，并据此进行定量预测。

(2)用前期海温特征进行预报。异常海温的分布具有范围大、厚度深、持续时间长等特点。大气与海洋是相互作用的，在异常大气环流出现之前往往早就出现海温异常，从而为长期水文预报提供了信息。与应用前期环流形势预报类似，在做法上也是分析历史海温资料与对应的预报对象的关系，概括出旱涝年前期海温分布的模式，由前期海温作出后期水文情况的定性预测；或考虑海温在时间上与空间上的连续性，在关键时段内挑选若干个点的海温作为预报因子，与预报对象建立回归方程并进行定量预报。

(3)用太阳活动规律进行预报。太阳辐射是供给大气运动的根本能量来源。太阳活动增强时，太阳紫外线辐射、微粒辐射以及无线电范围的辐射都有增强，从而引起大气运动状态的改变而影响到水文过程。太阳黑子相对数的多少基本上反映了太阳活动的强弱。黑子相对数有着平均为 11 年等的多种周期。许多研究结果表明，不少地区的旱涝灾害往往发生在黑子相对数 11 年周期的极值点附近。因此，目前在这方面的预报方法主要是根据黑子数在 11 年周期中的位相或分析黑子数与江河水量变化之间的关系对后期可能发生的旱涝进行定性预测。

(4)用其他天文地球物理因素进行预报。一些研究结果表明，地球自转速度的变化，行星运动的位置，火山爆发，臭氧的多少等等对大气运动与水文过程都有一定的影响。分析了这

些因素与水文过程的对应关系后，可由此对后期可能发生的水文情况作出定性预估。

(5)用概率统计方法进行预报(简称统计预报)。用概率统计方法预报的基本原则是从大量历史资料中应用数理统计方法去寻找分析水文要素历史变化的统计规律以及与其他因素的关系，然后应用这些规律来进行预报。在长期预报中应用概率统计方法虽早已开始，但在20世纪的前半叶，研究长期预报的工具几乎局限于一般的相关、回归分析。最近二三十年来，由于随机过程、信息论、多元分析技术、时间序列分析以及计算技术的迅速发展与推广，使统计预报得到了很大的发展。目前统计预报的方法虽然很多，但大体上可归纳为两大类。

①多元分析。把江河水量等预报对象作为随机变量，把各个影响因素作为预报因子，然后应用多元分析的方法，如逐步回归、判别分析等对其进行筛选，建立预报方程进行定量预测。此外，在预报中为了达到客观分型、浓缩信息、简化计算等目的，经常应用聚类分析、主分量(自然正交函数)分析、典型相关等方法来作为数据处理的手段。

②时间序列分析。把预报对象作为一个离散化的平稳随机过程，应用线性自回归即AR(p)模型(p为模型阶数)进行预报。考虑到水文时间序列的非平稳性，60年代前后主要采取把序列分解为趋势项、周期项、平稳项、噪声项，然后分项预报，再进行叠加来获得预报结果。70年代以后，自回归求和模型，自回归移动平均，即ARMA(p,q)模型，自回归移动平均求和即AR IMA(p,d,q)模型以及处理非线性问题的门限自回归(TAR)模型均已应用于径流的长期预报。

二、中期预报方法

50年代应用高空气象资料直接预报洪水的中期预报方法至今仍在应用。但是，随着气象资料系列的增长与完整，对引起当地暴雨天气形势的分析也更为深入、细致，谐波分析的方法在中期预报中已得到应用。此外，时间序列分析等统计预报方法也已应用在中期预报上。

第二章　预报方法之一　历史演变与周期均值叠加

第一节　历史演变法

一、方法的基本思路

历史演变法是杨鉴初于1951年提出制作气象要素长期预报的一个方法。1958年由内蒙古自治区水文总站用来对黄河进行长期洪水预报。历史演变法是利用某一测站某一水文要素的历史演变曲线(过程线)的外形特征来制作预报的。它的基本出发点是：任何一个水文要素的长期记录反映了这一要素全面的历史变化。尽管影响这一要素的外部因素与内部因素十分复杂，直至目前还不能把它们一一辨认出来，也不能一一确定各个因素的影响程度，但这些因素的综合影响却都已毫无遗漏地反映在这个要素的历史记录之中。因此，只要能够找出这一要素的演变规律，就可利用这些规律来进行预报。

在实际工作时，根据历史演变曲线的情况，可归纳出以下几种规律。

(1)持续性。所谓持续性是指水文要素的历史变化中升降变化的持久程度。以头屯河哈地坡水文站年平均流量的历史演变曲线为例(见图 2-1)。

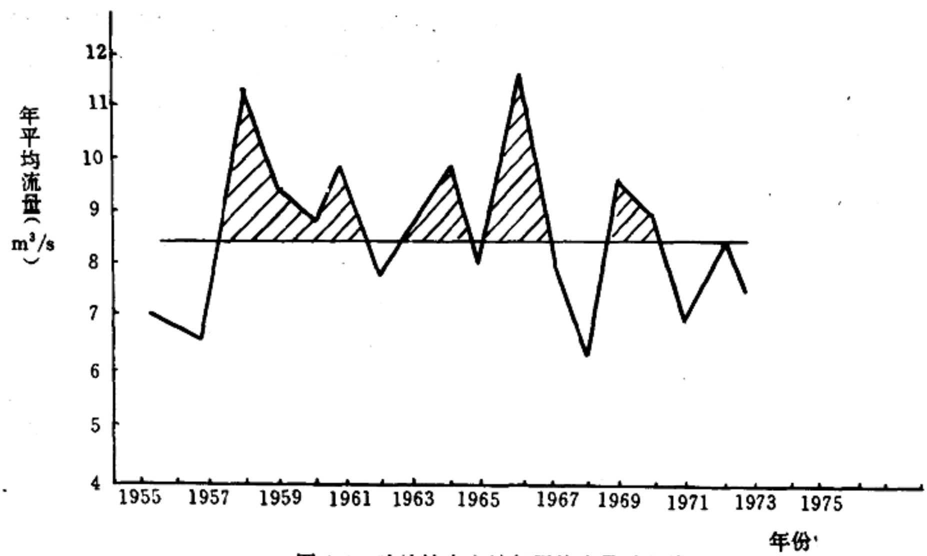

图 2-1　哈地坡水文站年平均流量过程线

注：本例资料年代嫌短，主要用来说明方法的基本思路。

①在 1955 年至 1973 年期间，如果某年年平均流量上升，并且超过 $8m^3/s$ 时，则第二年必然下降。在图 2-1 中发生 5 次这样的情况，无一例外。

②如果第一年下降，第二年继续下降，则第三年必然上升。在上述年代里，发生过 3 次这样的情况，也无一例外。

(2)相似性。相似性是指历史变化曲线在某一段时期与另外一段时期的变化趋势，在外形上有很相似的地方。例如在图 2-1 中可看到从 1957 年～1962 年与 1965 年～1970 年、1968 年～1973 年演变趋势非常相似。都表现为升—降—降—升—降。而且都是前一个峰点比后一个峰点高。

(3)周期性。这里所指的周期性并不是严格的周期，而是指某一水文要素的历史变化经过一定的时间间隔以后出现类似的变化。同时周期性本身在历史变化中也不是永远不变的，需要适当掌握，并配合其他规律来进行预报。如图 2-1 中从 1962 年至 1971 年期间每隔 3 年就出现一个低谷，它的数值均在均值以下。又如从 1958 年至 1964 年、1966 年至 1972 年期间每隔 3 年就出现一个峰点，除 1972 年的峰值稍低于均值外，其余峰值都在均值以上。因此，哈地坡站的 3 年周期性是比较明显的。

(4)最小可能性与最大可能性。任何水文要素的历史变化，它的数值在一定时期内均有适当的变化范围，注意变化范围对预报有很大的帮助。一个水文要素的数值超出它历史变化的范围不是不可能的，但是这种可能性较小。历史记录年代越长，超出这种范围的可能性也越小。因此，称超出历史变化范围的情况为最小可能性，而历史变化中经常出现的范围就叫最大可能性。如图 2-1 中哈地坡站年平均流量数值大于 $10m^3/s$ 的是最小可能性，同样低于 $6m^3/s$ 的也是最小可能性。年平均流量历史变化范围约在 6～10 之间，最大可能性在 7～9 之间。

(5)转折点。转折点是指历史变化中在某一时期内很明显的特征，在后一时期中有所改变，并出现新的特征，称这两个时期之间的转折期为转折点。因此，转折点就是两种历史变化时期中的分界线。在历史变化中转折点不是经常出现的，如图 2-1 中从 1955 年～1962 年历史变化曲线的趋势都是降—降—升，但是到 1962 年时变为降—升—升，因此 1962 年可以看作为一个转折点。

二、分析与预报的步骤

上面已阐述了水文要素历史变化的一般规律，现介绍怎样分析历史变化与作出预报的步骤。分析历史变化，记录年份至少需要二三十年以上，如果记录年份过少，则预测的可靠程度将降低，应用中可能会出现问题。分析和预报的方法有以下几个步骤。

(1)整理记录。首先按照预报要求，将资料加以整理和计算，然后画出历史演变曲线。横坐标表示年份，纵坐标表示要素的实际数值或距平值。用实际数值时，同时要计算历年的平均值，用横线画在历史演变曲线上，以便看出与均值的关系。

(2)分析主要规律。根据上述五种规律，检查一下那一种规律比较明显。计算每种规律在全部时间内出现的次数，例外的情况能否用其他规律补救，转折点的年份有什么特征等。把这些结果一一记下，进行综合研究与判断。

(3)做出初步预报值。应用上一步骤中已发现的主要规律，初步推断出明年的数值，这时特别要注意近几年的变化与全部历史变化曲线内哪些时期情况相似，尤其要注意上一年到本年的情况有什么变化。

(4)校核预报值。初步预报做完后，再把最后三五年包括下一年预报值在内的演变趋势和全部历史变化曲线各时间比较一下：是否相似的次数很多，例外的很少；预报值是否在最大可能范围以内，如果预报值太高或太低是否有充分的理由，这些理由是不是完全符合已有

资料的历史变化规律。

例如对哈地坡站作1974年年平均流量预报时，首先考虑它升1年降2年的持续性以及3年出现谷点的周期性甚为明显，因此，报1974年年平均流量将继续下降。至于下降到哪一点，则根据最近几年，即1966年～1968年，1969年～1971年的两个3年周期内第二年下降的幅度都在1.5m^3/s左右，但根据1973年6.76m^3/s的数值再下降1.5m^3/s时，数值为5.26m^3/s，超出了历史变化曲线的范围。因此，大体上可报在6.0m^3/s左右（实况是4.58m^3/s）。

综上所述，历史演变法提出了分析要素演变曲线变化规律的五个方面，是比较全面的，而且方法简便易行。其缺点是缺乏严格的定量标准，因此，在作预报时可能会因人而异。此外，还缺乏与影响水文状况的各个因素进行多方面的比较等。但尽管如此，由于上述优点，在长期预报已有不少进展的今天，历史演变法仍然没有失去它的现实意义。

第二节　周期均值叠加

如上节所述，应用“历史演变法”作预报时缺乏定量的形式并且包含一定主观成分。这就提出了一个问题，即能否根据过去的历史资料分析它们的统计规律作出未来的定量预报，概率统计方法在这方面提供了若干途径。本节介绍利用周期均值外推后叠加作出预报的方法，简称周期均值叠加。

一、方法的基本思路

一个水文要素随时间变化的过程尽管多种多样，但总可以把它看成是有限个具有不同周期的周期波互相重叠而形成的过程，其数学模型为

$$x(t) = p_1(t) + p_2(t) + \cdots + p_n(t) + \varepsilon(t) = \sum_{i=1}^{n} p_i(t) + \varepsilon(t) \tag{2-1}$$

式中，$x(t)$为水文要素序列，$p_i(t)$为第i个周期波序列，$\varepsilon(t)$为误差项。只要根据实测的水文要素数据，分析识别出水文要素所含有的周期，而且这些周期在预测区间内仍然保持不变的话，那末就可以根据分析出来的周期分别进行外延，然后再叠加起来进行预报，这种方法称为周期叠加。显然这个方法的关键问题是如何对实测数据进行周期分析的问题。还需要指出的是：由于影响水文要素长期变化因素的复杂性，这里所指的周期不可能象天体运动、潮汐现象所具有的严格周期，而只是概率意义上的周期。也就是只能理解为某一水文现象出现之后，经过一定的时间间隔，这种现象再次重复出现的可能性较大而已。

由于各种水文要素过程线的外形比较复杂，在图形上不易直接判断它是否存在某种周期，因此，需要借助于某些数理统计方法来对它进行识别。周期分析应该解决的问题是：

(1)某一水文要素是否存在周期；

(2)如果存在周期，周期是多长；

(3)根据实测数据分析得到的周期，它的可靠性如何。

分析周期的方法很多，在这一节里主要介绍应用方差分析识别周期的概念和计算方法。

二、应用方差分析识别周期的概念

在介绍方差分析之前,先看一个假想具有5年严格周期的特例。设某一水文要素具有20年的观测资料。其数据如下：

年份	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962
数据	10.0	8.0	7.0	6.0	5.0	10.0	8.0	7.0	6.0	5.0
年份	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
数据	10.0	8.0	7.0	6.0	5.0	10.0	8.0	7.0	6.0	5.0

如果按其存在的周期分组排列成表2-1时,由表可见,各组组内的数据都一样,没有任何差异,而各组的组平均值之间却差异较大。

表2-1　数据按5年周期排列表

项　数	一	二	三	四	五
1	10.0	8.0	7.0	6.0	5.0
2	10.0	8.0	7.0	6.0	5.0
3	10.0	8.0	7.0	6.0	5.0
4	10.0	8.0	7.0	6.0	5.0
组合计	40.0	32.0	28.0	24.0	20.0
组平均	10.0	8.0	7.0	6.0	5.0

如果不按它存在的周期排列,而按另一种时间间隔(比如4年)排列成表2-2时,则可以看到各组组平均值完全相同,没有任何差异,而各组内部的数据却差异较大。

表2-2　数据按4年周期排列表

项　数	组别			
	一	二	三	四
1	10.0	8.0	7.0	6.0
2	5.0	10.0	8.0	7.0
3	6.0	5.0	10.0	8.0
4	7.0	6.0	5.0	10.0
5	8.0	7.0	6.0	5.0
组合计	36.0	36.0	36.0	36.0
组平均	7.2	7.2	7.2	7.2

一般情况而言,如果这一要素存在着周期性变化,那末按其存在周期分组排列时,处在同一组内的各个数据应该是在同一位相之下的观测值,不同组的数据则为不同位相下的观测值。在各个周期高峰时期的观测值平均来说是比较大的,而在各低谷时期的观测值则是比较小的,峰谷之间的转换时期是中等的。因此,在同组之内的数据差异应该相对较小,而组与组之间的数据差异应该较大。

反之,如果不按它的周期分组列表时,则同组之内的数据包含着不同位相之下的观测

值，即一组之内的数据可能既有高峰，又有低谷和中间数据值。这样就造成同组之内的数据相差较大，而组与组之间的数据相差较小的现象。

组间与组内数据差异的情况，可分别用计算离差平方和的方法把它反映出来，并且把它们进行比较，从而推断是否有周期存在，这就是利用方差分析作周期分析的基本概念。

三、组间离差平方和与组内离差平方和的计算

若把 n 个观测数据 $x_1, x_2, \cdots, x_n$ 排列成表 2-3 的形式，其中 j 表示组别，$j=1,2,\cdots,b$ 表示分为 b 组。i 为每组含有的项数，$i=1,2,\cdots,a$ 表示每组有 a 个数据，T_j 为每组的合计数。T 为总数，$\overline{x}_j$ 为每组的组平均值，组间离差平方和与组内离差平方和的计算公式如下：

组间离差平方和

$$S_1 = a\sum_{j=1}^{b}(\overline{x}_j - \overline{x})^2$$

组内离差平方和

$$S_2 = \sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij} - \overline{x}_j)^2$$

表 2-3　试验周期分组排列表

i		试验周期分组 (j)			
		1	2	…	b
每组项数	1	x_{11}	x_{12}	…	x_{1b}
	2	x_{21}	x_{22}	…	x_{2b}
	⋮	⋮	⋮	⋮	⋮
	a	x_{a1}	x_{a2}	…	x_{ab}
T_i		T_1	T_2	…	T_b
x_j		x_1	x_2	…	x_b
T_j^2		T_1^2	T_2^2	…	T_b^2
T_j^2/a_j		T_1^2/a_j	T_2^2/a_j	…	T_b^2/a_j

直观地讲，$\sum_{j=1}^{b}(\overline{x}_j-\overline{x})^2$ 表示了各组组平均值 $\overline{x}_j$ 对总平均值 $\overline{x}$ 的离差平方和。$\sum_{j=1}^{b}(\overline{x}_j-\overline{x})^2$ 前面再乘上一个 a，这是因为考虑到每组内又有 a 项的关系，它表达了组与组之间数据差异的情况。组内离差平方和是指同组之内各个数据 x_{ij} 对组平均值 $\overline{x}_j$ 的离差平方再求和，即 $\sum_{i=1}^{a}(x_{ij}-\overline{x}_j)^2$，代表了每组内部数据的差异情况。由于共有 b 组，所以还要把各组的 $\sum_{i=1}^{a}(x_{ij}-\overline{x}_j)^2$ 相加。得

$$S_2 = \sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij} - \overline{x}_j)^2$$

公式来源证明如下：

设将 n 个观测数据排列成表 2-3 的形式，则这一组数据总的离差平方和应为 $S=\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x})^2$，现将它进行分解为组间离差平方和与组内离差平方和之和，为

$$\begin{aligned}\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x})^2 &= \sum_{j=1}^{b}\sum_{i=1}^{a}[(x_{ij}-\bar{x}_j)+(\bar{x}_j-\bar{x})]^2 \\ &= \sum_{j=1}^{b}\sum_{i=1}^{a}[(x_{ij}-\bar{x}_j)^2+2(x_{ij}-\bar{x}_j)(\bar{x}_j-\bar{x})+(\bar{x}_j-\bar{x})^2] \\ &= \sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x}_j)^2+2\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x}_j)(\bar{x}_j-\bar{x}) \\ &\quad+\sum_{j=1}^{b}\sum_{i=1}^{a}(\bar{x}_j-\bar{x})^2 \\ &= \sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x}_j)^2+a\sum_{j=1}^{b}(\bar{x}_j-\bar{x})^2 \\ &\quad+2\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x}_j)(\bar{x}_j-\bar{x})\end{aligned}$$

其中

$$\begin{aligned}2\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x}_j)(\bar{x}_j-\bar{x}) &= 2\sum_{j=1}^{b}[(\bar{x}_j-\bar{x})\sum_{i=1}^{a}(x_{ij}-\bar{x}_j)] \\ &= 2\sum_{j=1}^{b}[(\bar{x}_j-\bar{x})(\sum_{i=1}^{a}x_{ij}-\sum_{i=1}^{a}\bar{x}_j)] \\ &= 2\sum_{j=1}^{b}[(\bar{x}_j-\bar{x})(T_j-a\bar{x}_j)] \\ &= 2\sum_{j=1}^{b}[(\bar{x}_j-\bar{x})(T_j-T_j)] \\ &= 0\end{aligned}$$

故

$$\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x})^2 = a\sum_{j=1}^{b}(\bar{x}_j-\bar{x})^2+\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij}-\bar{x}_j)^2 \tag{2-2}$$

在实际计算时，由于列表计算的方便以及为了减少计算误差，一般采用下面的公式进行计算：

$$S_1=\sum_{j=1}^{b}\frac{T_j^2}{a}-\frac{T^2}{n} \tag{2-3}$$

$$S_2=\sum_{j=1}^{b}\sum_{i=1}^{a}x_{ij}^2-\sum_{j=1}^{b}\frac{T_j^2}{a} \tag{2-4}$$

证明如下：

组间离差平方和，为

$$S_1=a\sum_{j=1}^{b}(\bar{x}_j-\bar{x})^2$$

$$
\begin{aligned}
&= a\sum_{j=1}^{b}\bar{x}_j^2 - 2a\sum_{j=1}^{b}\bar{x}_j\bar{x} + a\sum_{j=1}^{b}\bar{x}^2 \\
&= a\sum_{j=1}^{b}\bar{x}_j^2 - 2a\bar{x}\sum_{j=1}^{b}\bar{x}_j + ab\bar{x}^2 \\
S_1 &= a\sum_{j=1}^{b}\bar{x}_j^2 - 2ab\bar{x}^2 + ab\bar{x}^2 \\
&= a\sum_{j=1}^{b}(\frac{T_j}{a})^2 - \frac{a^2b^2}{ab}\bar{x}^2 \\
&= \sum_{j=1}^{b}\frac{T_j^2}{a} - \frac{1}{n}(a\sum_{j=1}^{b}\bar{x}_j)^2 \\
&= \sum_{j=1}^{b}\frac{T_j^2}{a} - \frac{1}{n}(\sum_{j=1}^{b}T_j)^2 \\
&= \sum_{j=1}^{b}\frac{T_j^2}{a} - \frac{T^2}{n}
\end{aligned}
\tag{2-5}
$$

由于在排表时，数据不一定恰好排满表格，即无共同的项数 a，此时 a 改为 a_j，则：

$$
S_1 = \sum_{j=1}^{b}\frac{T_j^2}{a_j} - \frac{T^2}{n} \tag{2-6}
$$

组内离差平方和：

$$
\begin{aligned}
S_2 &= \sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij} - \bar{x}_j)^2 \\
&= \sum_{j=1}^{b}\sum_{i=1}^{a}x_{ij}^2 - 2\sum_{j=1}^{b}\sum_{i=1}^{a}x_{ij}\bar{x}_j + \sum_{j=1}^{b}\sum_{i=1}^{a}\bar{x}_j^2 \\
&= \sum_{j=1}^{b}\sum_{i=1}^{a}x_{ij}^2 - 2\sum_{j=1}^{b}(\bar{x}_j\sum_{i=1}^{a}x_{ij}) + a\sum_{j=1}^{b}\bar{x}_j^2 \\
&= \sum_{j=1}^{b}\sum_{i=1}^{a}x_{ij}^2 - 2\sum_{j=1}^{b}(\frac{T_j}{a}\cdot T_j) + a\sum_{j=1}^{b}(\frac{T_j}{a})^2 \\
&= \sum_{j=1}^{b}\sum_{i=1}^{a}x_{ij}^2 - \sum_{j=1}^{b}\frac{T_j^2}{a}
\end{aligned}
\tag{2-7}
$$

同样，当无共同的 a 时，则以 a_j 代替：

$$
S_2 = \sum_{j=1}^{b}\sum_{i=1}^{a}x_{ij}^2 - \sum_{j=1}^{b}\frac{T_j^2}{a_j} \tag{2-8}
$$

四、方差比 F 的确定与 F 检验

在周期分析时，由于事先不能知道周期数目，因此需要从可能存在的周期中反复排列各种数值的周期试验表，同时计算各种试验周期的 S_1 与 S_2，然后分析比较，决定存在哪种周期。由于不同的试验周期在分组数目与每组含有的项数上都有不同，因此为了能够互相比较，还要计算它们的平均情况，即平均组间离差平方和（组间方差）与平均组内离差平方和（组内方差）。它们的计算公式如下：

$$
组间方差 = \frac{S_1}{f_1}
$$

$$组内方差=\frac{S_2}{f_2}$$

上式中，f_1，f_2 分别为组间离差平方和与组内离差平方和的自由度。自由度的意义可理解为：如果要获得一个系列 $x_1,x_2,\cdots,x_n$，倘若对这个系列没有任何附加条件，则系列中的 n 个变数可以自由选择，不受任何限制，在这种情况下，称它具有 n 个自由度，即 $f=n$。如果对这个系列加了一个限制条件，比如，要这个系列的平均值等于 $\bar{x}$，那末系列的 n 个变数中，只有 $n-1$ 个可以自由选择，而其余一个由于 $\bar{x}$ 条件的限制，不能再自由选取。在这种情况下就只具有 $n-1$ 个自由度了，即 $f=n-1$。同样，如果有 k 个限制条件，则自由度 $f=n-k$。组间离差平方和的自由度 f_1 是由 b 个组的组平均值的离差平方和构成的，当其中任意 $b-1$ 项确定之后，最后一项就不能再任意变动，所以，组间离差平方和的自由度 $f_1=b-1$。组内离差平方和的自由度是由 b 个组的组内离差平方和构成的，而每个组又由 a 项构成。每组的自由度为 $a-1$，现共有 b 个组，故组内离差平方和的自由度 $f_2=b(a-1)=(b\times a)-b=n-b$。这组资料总的离差平方和的自由度显而易见为 $f=n-1$，故有 $f=f_1+f_2$，称为自由度的分解。

有了组间方差与组内方差之后，就可对各个试验周期的组间、组内数据离散情况进行比较了，但是要达到什么标准才算组间方差显著地大于组内方差，这需要计算它们之间的比值来确定，即方差比 F 为

$$F=\frac{\frac{S_1}{f_1}}{\frac{S_2}{f_2}}=\frac{\dfrac{\sum_{j=1}^{b}\frac{T_j^2}{a_j}-\frac{T^2}{n}}{b-1}}{\dfrac{\sum_{j=1}^{b}\sum_{i=1}^{a}x_{ij}^2-\sum_{j=1}^{b}\frac{T_j^2}{a_j}}{n-b}} \tag{2-9}$$

在一定条件下，可以证明方差比 F 是一个随机变量，而且是服从 F 分布的。F 分布有专用的 F 分布表可查(见表 2-4，表 2-5，表 2-6)。因此，可以应用 F 检验的方法检验组间方差是否显著地大于组内方差(详细的推导可参看周华章编的《工业技术应用数理统计学》下册)。

在实际应用时，先根据实测资料计算出方差比 F 的数值，然后由计算得到的自由度 f_1，f_2 与选定的信度 α(信度的意义可理解为发生判断错误可能性的概率，如 $\alpha=0.05$，表示发生判断错误可能性的概率为 5%，也可以说判断正确的保证率为 95%)。在 F 分布表中查出相应的 F_α，如果

$F>F_\alpha$ 则表明在这一信度水平上，差异显著，有周期存在。

$F\leqslant F_\alpha$ 则表明在这一信度水平上，差异不显著，无周期存在。

在分析水文要素的数据是否存在着周期时，可根据数据的数目 n，列出可能存在的周期，一般可能存在的周期为 $2,3,\cdots,k$。

$$k=\left[\frac{n}{2}\right]=\begin{cases}\frac{n}{2} & n\text{ 为偶数}\\ \frac{n-1}{2} & n\text{ 为奇数}\end{cases}$$

然后按 $2,3,\cdots,k$ 的数值，分别排表计算它们的方差比 F，在这些 F 值中挑选最大的 F 值与选定信度下的 F 临界值 F_α 作比较分析，决定是否存在周期和存在何种周期。下面以实

例说明计算步骤。

五、数值例子

以浙江省钱塘江流域金华江金华水文站年最高水位的预报说明本方法的计算过程。金华站自 1952 年开始有水位记录，1960 年站址迁移至河畔桥，根据两处落差关系统一转换至河畔桥断面处的水位，至 1973 年共有 22 年资料，全部记录见表 2-7 第(1)列。

金华站年最高水位长期预报的基本思路是把逐年年最高水位的变化过程作为一个波动过程看待，并认为这个波动是由 n 个具有不同周期的周期波叠加形成的。其数学模型为

$$x(t)=\sum_{i=1}^{n}p_i(t)+\epsilon(t)$$

式中，$x(t)$为年最高水位序列，$p_i(t)$为各个周期波序列，$\epsilon(t)$为误差项序列。用方差分析识别周期的计算步骤如下。

(1)计算各年的年最高水位距平值 ΔH_t。

$$\Delta H_t=H_t-\overline{H}$$

式中，$\overline{H}$ 为历年年平均最高水位，H_t 为各年实测的最高水位。金华站自 1952 年至 1973 年的年平均最高水位为 37.81m，1952 年最高水位为 38.71m，故 1952 年的距平值为 38.71－37.81＝0.90m。计算结果见表 2-7 第(2)列。并以各年的距平值组成一个时间序列，记以 $x(t)$。

当手算时，利用距平进行的好处是：

①由于减去了一个大的数值，因此，可以减少计算工作量及计算差错。

②由于$\frac{T^2}{n}\to 0$，故组间离差平方和的计算公式可简化为

$$S_1=\sum_{j=1}^{b}\frac{T_j^2}{a_j}$$

(2)将各年的距平值按各种试验周期进行分组排表。金华站共有 22 年资料，n 为偶数，故 $k=\frac{22}{2}=11$。

因此，试验周期应从 11 年排至 2 年，现将金华站 7 年试验周期的排表情况列在表 2-8 中。

表 2-4　F 分布表($\alpha=0.01$)

f_2	f_1 = 1	2	3	4	5	6	7	8	9	10	12	15	20	60	∞
1	4052	5000	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6313	6360
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.48	99.50
3	34.12	30.82	29.46	28.71	23.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.32	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.65	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.20	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.06	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	5.82	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.03	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.48	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.08	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	3.78	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.54	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.34	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.18	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.05	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	2.93	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	2.83	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	2.75	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.67	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.61	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.55	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.50	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.45	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.40	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.36	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.21	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.02	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	1.84	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.66	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.47	1.00