

随机水文学(三)

邓育仁 高荣松 丁晶

(成都科技大学)

第二讲 水文系列分析

对水文变量(如江河的流量,湖泊及井中水位等)进行观测,得到以时间为自变量(一般取等距时间)的系列观测值,称为水文时间系列(简称水文系列)。例如,年平均流量系列等。

我们把水文变量随时间而变化的系列观测值看作水文随机过程的一个现实,通过对它的统计分析,估计水文随机过程的总体特性和建立最佳水文随机模型,以便于各方面的实际应用。

本讲从组成水文系列的成分出发,分别讲述各种成分的概念、统计特征以及识别和描述方法。有关的其他内容,放在以后有关讲中介绍。

一、水文系列组成

一般认为水文系列 x_t 是由趋势成分 T_t 、周期成分 P_t 、随机相依成分 η_t 和纯随机成分 ε_t 线性叠加组成,即

$$x_t = T_t + P_t + \eta_t + \varepsilon_t \quad t=1, 2, \dots \quad (2 \cdot 1)$$

式中,右端前两项称为确定性成分,后两项称为随机性成分(用 ξ_t 表示)则有

$$\xi_t = \eta_t + \varepsilon_t$$

由式(2.1),当 $T_t + P_t + \eta_t \rightarrow 0$ 时,水文系列仅由单一的纯随机成分组成,即

$$x_t = \varepsilon_t$$

如年最大流量系列等,就是一个纯随机系列。

由于水文系列多是复杂的合成系列,因此如何把合成系列中的各种成分划分出来,并用适当的数学模型进行描述,这是研究水文系列性质的重要内容,也是水文系列模拟的主要方法之一。

以上所称系列组成的“成分”,也称为“项”、“分量”及“系列”。以下将不加区别的使用这些同义词。

下面我们分别讲述水文系列组成中的各种成分。

二、趋势成分

(一)一般概念

随着时间增长，水文系列中的各值平均或是增加或是减少，这种有规律地变化，称为趋势（见图 2.1）。

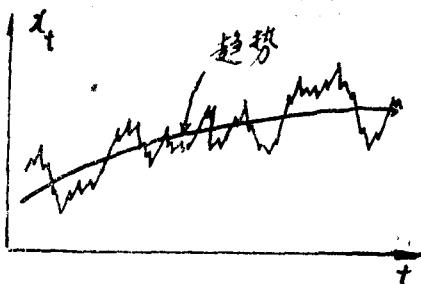


图 2.1 非线性趋势成分

趋势是水文系列中的暂态成分，常被叠加在其他成分中。实际工作中，要求水文系列具有原始状态或一致性，若系列中包括暂态成分就破坏了这种性质，利用这种系列预估未来事件，可能被大大地歪曲。因此要求识别（查明）和描述这种成分，并设法从系列中排除。

趋势可能是线性的，也可能是非线性的。描述趋势常用多项式

$$T_t = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n \quad (2.2)$$

式中 $a_0, a_1, a_2, \dots, a_n$ 为系数。

一般多用简单形式

$$T_t = a_0 + a_1 t \quad (2.3)$$

引起趋势的原因有人为的因素和自然的因素。如气候因素的明显变化，引起径流有一定趋势地变化；沿河不断筑堤防洪，年最大洪峰流量系列有增大的趋势；流域内灌溉工程不断兴建，流域总蒸发量有增加趋势，而径流量则有减少的趋势等。

(二) 趋势成分的检验、描述和排除

1. 滑动平均法 按水文系列的几个前期值和后期值取平均或取 $2k$ 或 $2k+1$ 个相连值的平均，给出变量值 y_t ，使系列光滑化：

$$y_t = \frac{1}{2k+1} \sum_{-k}^k x_{t+i} \quad (2.4)$$

选择合适的 k ，使系列高频的振荡平均掉，但不能过分。

滑动平均光滑后的系列 y_t ，若显示出趋势，就可从原来的系列中排除掉。

2. 线性趋势的回归检验 如果趋势近于线性，则用线性回归进行检验。由式 (2.3) 和 (2.1) 有

$$x_t = a_0 + a_1 t + u_t \quad (2.5)$$

式中 $u_t = P_t + \xi_t$ 称为剩余（或残差）成分。用回归分析方法求出系数 a_0 和 a_1 的估计值 \hat{a}_0 和 \hat{a}_1 ，再进行回归效果检验。当在显著水平 α 下，拒绝原假设 $a_1 = 0$ 时，回归效果显著，即线性趋势明显；相反，线性趋势则不明显。

检验分析趋势成分的方法还有线性时相关法，游程检验法，Kendall 序次相关检验等。

三、自相关分析和谱分析

自相关分析和谱分析是研究水文系列统计性质的有利工具。在讨论周期成分和随机成分之前，应先了解这两种分析方法。

(一) 自相关分析

自相关分析用来决定以推移时间 k 隔开的系列连贯值之间的线性相依性质。通常用自相关系数和自相关图研究。

由实测水文系列按下式计算自相关系数：

$$\gamma(k) = \frac{C_{ov}(x_t, x_{t+k})}{(V_{av}x_t \cdot V_{av}x_{t+k})^{1/2}} \quad (2 \cdot 6)$$

其中协方差函数

$$C_{ov}(x_t, x_{t+k}) = \frac{1}{N-k} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}_{t+k})$$

方差

$$V_{av}x_t = \frac{1}{N-k} \sum_{t=1}^{N-k} (x_t - \bar{x})^2$$

$$V_{av}x_{t+k} = \frac{1}{N-k} \sum_{t=1}^{N-k} (x_{t+k} - \bar{x}_{t+k})^2$$

均值

$$\bar{x}_t = \frac{1}{N-k} \sum_{t=1}^{N-k} x_t$$

$$\bar{x}_{t+k} = \frac{1}{N-k} \sum_{t=1}^{N-k} x_{t+k}$$

在大样本(一般指样本容量在 50 以上)时, x_t, x_{t+k} 可用样本均值 $\bar{x} = \frac{1}{N} \sum_{t=1}^N x_t$ 代替,

则用下式计算：

$$\gamma(k) = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2} \quad (2 \cdot 7)$$

式(2·6、7)中 N 为系列长度; $k = 0, 1, 2, \dots, m$; 当 $N > 50$ 时, m 取小于 $N/4$; 常取 m 在 $N/10$ 左右。

自相关系数 $\gamma(k)$ 随推移时间(简称时移或滞后) k 而变化的图形, 称为自相关图或相关图。

下面给出几种样本系列的相关图, 供后面分析使用。

1. 周期系列相关图 周期系列或隐含周期成分的系列, 其相关图也呈周期变化(见图2·2)。

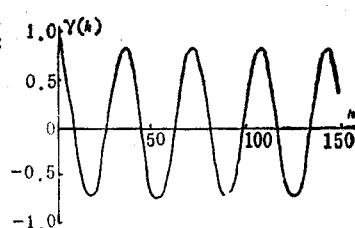


图 2·2 周期系列相关图 一个平稳的线性相依系列

的相关图是一条衰减的曲线，即随着时移 k 的增加， $\gamma(k)$ 的波动越来越小如(图2·3)。

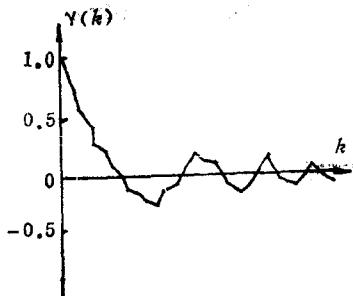


图 2.3 相依系列相关图

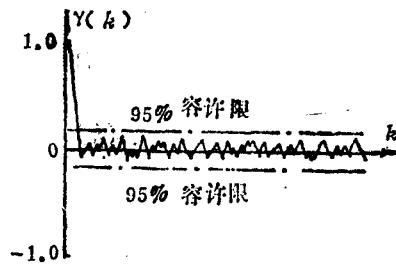


图 2.4 纯随机系列相关图

3. 纯随机系列相关图 纯随机系列相关系数围绕零，在容许限内作小的波动(见图2·4)。

(二) 谱分析

上述相关分析属于时间域分析。我们把水文系列视为一种有一定规律的振动现象，就可以认为它是由一组不同频率(或不同振幅或不同周期)的余弦波和正弦波组成的谐波叠加而成的，亦即可用傅立叶级数表示。同时又能通过对傅立叶级数的变换，得到系列的方差线谱或方差谱密度。若以这些谱为工具分解系列，找出对系列组成贡献大的谐波个数及其对应的频率(或周期)，这就有利于研究水文系列的内部结构。实际上谱分析的作用也象光谱分析一样(白光是一种合成光，它透过三棱镜时，分成红、橙、黄、绿、青、兰、紫七种不同振动频率的单一光)在频率域上分解水文系列，估计各谐波的频率(或周期)。

在频率域上常用方差线谱和方差谱密度分析水文系列，现分述如下。

1. 方差线谱 在一定条件下，可以用傅立叶级数近似描述水文系列，即

$$\left. \begin{aligned} x_t &= a_0 + \sum_{j=1}^k (a_j \cos \omega_j t + b_j \sin \omega_j t) \\ x_t &= a_0 + \sum_{j=1}^k A_j \cos(\omega_j t + \theta_j) \end{aligned} \right\} \quad (2 \cdot 8)$$

或

$$t = 1, 2, \dots, N$$

式中 系数 a_0, a_j, b_j 按傅立叶级数的有关公式计算；角频率 $\omega = \frac{2\pi}{N}$ ， N 为 系列长度； θ 为相位； k 为谐波个数；谐波振幅

$$A_j^2 = a_j^2 + b_j^2 \quad (2 \cdot 9)$$

称 $\frac{A_j}{2}$ 对 ω_j (或频率 $f_j = \frac{\omega_j}{2\pi}$) 关系为频谱图或周期图(如图2·5)。

可证明：

$$\sum_{j=1}^k \frac{A_j^2}{2} = \sigma^2$$

其中 σ^2 为系列方差，因此也把频谱图或周期图称为方差线谱。

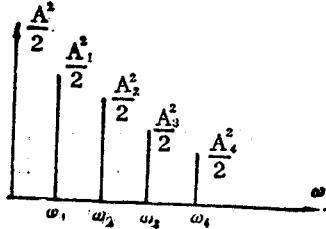


图 2.5 方差线谱(频谱)图

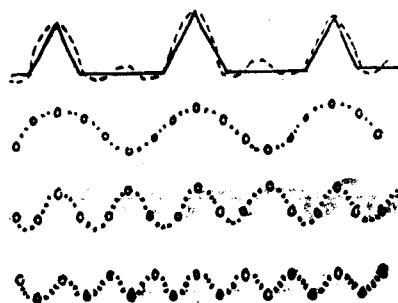


图 2.6 用有限个谐波和逼近周期函数的示意图

可取有限个贡献大的谐波之和的傅立叶级数去逼近某一个函数。如图 2.6 中的峰形函数可取三个(或更多个)主要谐波来逼近。方差线谱(如图 2.5)清楚表明在给定的函数中，所包含的谐波频率分量及各分量振幅(或周期)的大小。

2. 方差谱密度 同一般密度函数概念一样，角频率区间 $\Delta\omega$ (当 $\Delta\omega \rightarrow 0$)上的方差线谱与 $\Delta\omega$ 之比称为方差谱密度。显然，它的作用与方差线谱类似。

可证明由样本估计的方差谱密度 $S(\omega)$ (或 $S(f)$) 与样本相关系数 $\gamma(k)$ 有如下关系：

$$S(\omega) = \frac{1}{\pi} (1 + 2 \sum_{k=1}^m \gamma(k) \cos \omega k) \quad (2 \cdot 10)$$

或

$$S(f) = 2 [1 + 2 \sum_{k=1}^m \gamma(k) \cos 2\pi f k] \quad (2 \cdot 11)$$

式中符号意义同前。

由于按式(2·10、11)计算的方差谱密度离差较大，绘出的图形规律性差，常要进行平滑处理。平滑处理用到的函数称为谱窗或权函数。常用的有 Hamming 窗：

$$D(k) = 0.54 + 0.46 \cos(\pi k / m) \quad (2 \cdot 12)$$

和 Hann 窗：

$$D(k) = 0.50 + 0.50 \cos(\pi k / m) \quad (2 \cdot 13)$$

由式(2·12或13)代入式(2·11、12)得到经平滑处理的方差谱密度

$$S(\omega) = \frac{1}{\pi} [1 + 2 \sum_{k=1}^m D(k) \gamma \cos \omega k] \quad (2 \cdot 14)$$

或

$$S(f) = 2 [1 + 2 \sum_{k=1}^m D(k) \gamma(k) \cos 2\pi f k] \quad (2 \cdot 15)$$

这就是常用的计算公式。

3. 几种样本系列方差谱密度 (1) 周期系列方差谱密度。图 2·7 中，与

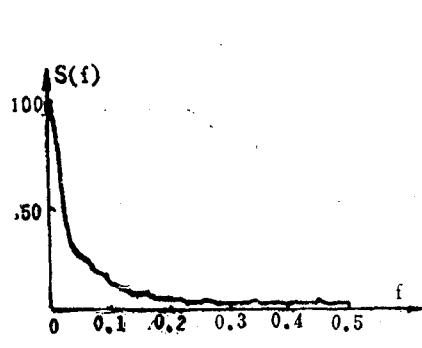


图 2.7 周期系列方差谱密度

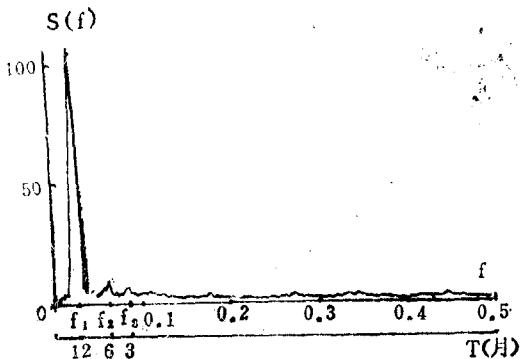


图 2.8 相依系列方差谱密度

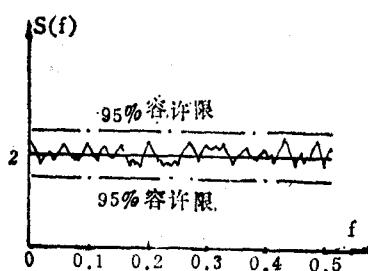


图 2.9 纯随机系列方差谱密度

(1) $S(f)$ 尖峰值对应的频率 (f_1, f_2, f_3) 或周期 (12月, 6月, 3月), 即为周期系列(或隐含周期成分的系列)中贡献最大的三个谐波的频率或周期。
 (2) 相依系列方差谱密度。平稳线性相依系列方差谱密度是一条总趋势为下降的曲线(如图 2.8)。
 (3) 纯随机系列方差谱密度。纯随机系列的方差谱密度 $S(f)$ 围绕常数 2 在容许限内波动(如图 2.9)。

上述相关分析和谱分析在数学上是等价的, 它们在水文系列分析中都有广泛应用。

四、周期成分

(一) 一般概念

水文系列中的确定性周期成分, 一种是由于地球绕太阳旋转和地球自转影响形成。前者如月径流量、月蒸发量等系列受季节影响, 明显存在以 12 个月为周期的周期成分。后者如气温及蒸发量等系列中存在以日为周期的周期成分。另一种是由于月球绕地球运转形成阴历月的周期。如河口潮汐水位系列, 就是受地球自转, 月球绕地球运转和地球绕太阳旋转三种循环的影响, 表现为近似周期的变化(以下不讨论这个问题)。

(二) 周期成分的识别和提取

将式(2.1)中趋势成分 T_t 排除后, 则有

$$x_t = P_t + \xi_t \quad (2.16)$$

设 $\tau = 1, 2, \dots, T$. T 是系列中基本周期; $t = 1, 2, \dots, n$; n 为 T 的数目(如 T 为 12 个月, n 为年数, $N = nT$ 是样本系列长度); $P_\tau = P_t$ 是均值中的周期成分。则式(2.16)写成

$$x_{t,\tau} = P_\tau + \xi_{t,\tau} \quad (2.17)$$

用傅立叶级数表示

$$P_t = \mu_x + \sum_{j=1}^d (a_j \cos \frac{2\pi j}{T} t + b_j \sin \frac{2\pi j}{T} t) \quad (2 \cdot 18)$$

其中 μ_x 为系列 x_t 的均值; d 是显著谐波分量个数; 系数按下式估计:

$$\left. \begin{aligned} a_j &= \frac{2}{nT} \sum_{l=1}^n \sum_{\tau=1}^T (x_{l,\tau} - \mu_x) \cos \frac{2\pi j}{T} \tau \\ b_j &= \frac{2}{nT} \sum_{l=1}^n \sum_{\tau=1}^T (x_{l,\tau} - \mu_x) \sin \frac{2\pi j}{T} \tau \end{aligned} \right\} \quad (2 \cdot 19)$$

由式 (2·19) 拟合周期成分, 必须先确定显著谐波个数 d 。一种粗略的识别方法就是利用样本方差谱密度 (图 2·7) 或方差线谱 (图 2·6) 直观地识别贡献大的谐波个数 d 及其各个的频率 (或周期)。

图 2·7 是大渡河铜街子站实测月径流系列方差谱密度, 可看出系列中隐含有贡献大的 3 个谐波, 其周期分别为 12 个月、6 个月和 3 个月。由此计算出周期成分

$$\begin{aligned} P_t &= 1485 - 796 \cos \frac{2\pi}{12} t - 1102 \sin \frac{2\pi}{12} t + 28 \cos \frac{2\pi}{6} t + 266 \sin \frac{2\pi}{6} t \\ &\quad + 89 \cos \frac{2\pi}{3} t + 140 \sin \frac{2\pi}{3} t \end{aligned}$$

五、相依成分

(一) 一般概念

由式 (2·1), 一般认为一个水文系列中, 当排除趋势成分、提取周期成分后, 剩余成分 ξ_t 就是一个平稳随机系列, 其中一部分是相依成分 (η_t), 另一部分是纯随机成分 (ε_t)。

一个水文系列, 如果其顺序各值之间的关系是一种线性的数学模型, 那么这种关系就是 (线性) 相依的。系列内部的这种相依关系, 其相依强弱通常用上述的相关系数作为定量的测度。

对于有的流域年径流系列, 若流域内含水层较厚, 河槽下切较深, 地面和地下调蓄能力较强, 有可能出现相邻年份 (甚至相隔年分) 之间年径流的关系比较密切, 反映出一阶相关系数 $\gamma(1)$ (甚至 $\gamma(2)$) 较大, 并且随着研究时段的缩短, 相依关系更为密切。如铜街子站年径流系列的 $\gamma(1)$ 为 0.43, 而月和旬径流系列的 $\gamma(1)$ 分别为 0.64 和 0.74。

(二) 相依成识别和拟合

相关图和方差谱密度是识别相依成分的有效工具。图 2·3、8 显示出平稳相依系列样本相关图和方差谱密度的特点, 据此识别样本系列的性质。还可以用相关系数检验法作进一步识别。

若系列相互独立地来自正态总体，它的理论相关系数

$$\rho(k) = \begin{cases} 1, & k=0 \\ 0, & k \neq 0 \end{cases}$$

当 $N-k$ 充分大时（如 $N-k > 50$ ），统计量

$$U(k) = \sqrt{N-k} \gamma(k) \quad (2 \cdot 20)$$

在 $k \neq 0$ 时，渐近服从标准正态分布。在给定显著水平 α 后，检验系列是相依的，且是独立的。

描述这类平稳随机系列的模型有自回归模型，滑动平均模型和自回归滑动平均模型等。这些模型将在第四讲介绍。

六、纯随机成分

系列任一时刻的值和任何其他时刻的值之间互不影响，则这种系列叫独立随机系列或纯随机系列。

纯随机系列理论相关系数

$$\rho(k) = \begin{cases} 1, & k=0 \\ 0, & k \neq 0 \end{cases}$$

从式 (2·10、11) 知，在理论上纯随机系列方差谱密度 $S(\omega) = \frac{1}{\pi}$ 或 $S(f) = 2$ 为常数。

纯随机系列样本相关图和方差谱密度分别围绕理论相关系数为零和方差谱密度为 $\frac{1}{\pi}$ （或 2）的上下波动（在给定容许限内），由此可识别样本系列的性质。

此外，式 (2·20) 也适用于纯随机系列的检验。

描述纯随机系列的概率模型常有正态分布型、对数正态分布型和皮尔逊Ⅲ型等。样本系列选择哪种模型描述，可以通过正态概率纸拟合检验和 Кочмогоров-Смирнов拟合优度检验来推断。但应指出，当样本容量不大时，用小样本推断总体的概率分布是很难得出正确结论的。对此，应十分重视样本系列物理性质分析和借鉴已有的经验，合理的选择概率模型。（待续）