3 MATHEMATICAL METHODS USED IN SYSTEMS THEORY

The mathematical methods used in the theory of systems are related to the nature and function of the systems analysed.

WRS have, in principle, a probability character. Therefore probability methods, and methods of mathematical statistics and the theory of stochastic processes are used in their investigation. The methods of Fourier, Laplace and z-transformation are used for the description of the dynamic complex behaviour of these systems, and the methods of the calculus of variations are applied for optimization in tasks where the structure and behaviour of systems are described and investigated.

A selection of the basic mathematical methods related to these issues is presented in this chapter; a more detailed treatment can be found in the specialist literature referred to in the text.

3.1 PROBABILITY THEORY

3.1.1 Basic Notions

Probability theory deals with investigation of *mass random events*; it is based on a large number of observations or measurements of phenomena occurring under a given set of conditions. For the determination of the statistical properties of events, many elementary events (viz. sample points) have to be determined (e.g. by measurement, observation etc.), if that is possible in the given case. In general, a large sample of these elements can be realized taken for random events from a sample space (e.g. possible outcomes of an experiment; all possible outcomes form the sample space). The occurrence of a random event can be predicted with some probability, but not exactly, even if the set of conditions governing its occurrence is maintained. If this probability is equal to one, the event is certain (sure), if it is equal to zero, the event is impossible.

The term probability can be more easily explained than defined. Often, the classical definition of probability is used. It says: If event A can be split into elementary events X_i $(i = 1, 2, \ldots, m)$ (i.e. events that cannot be further split into particular events), the probability of its occurrence in the given complete system of n mutually exclusive and equally probable events (where the sum of all the elementary events is a certain event and each pair of elementary events has a zero product) is

$$P(A) = \frac{m}{n} \qquad (m \leqq n) \tag{3.1}$$

In the case of a complete system of n equally probable events $A_1, A_2, \ldots, A_n$ which the event A is composed of and if

$$P(A) = \sum_{i=1}^{n} P(A_i) \tag{3.2}$$

then

$$P(A) = \sum_{i=1}^{n} p = np = 1 \tag{3.3}$$

so that

$$p = \frac{1}{n} \tag{3.4}$$

The *unconditional probability* of occurrence of event A was related only to a system of conditions. If event A also depends on the occurrence of an other event, B (given that the event B has occurred), then

$$P_B(A) = \frac{P(AB)}{P(B)} = \frac{P(A)\,P_A(B)}{P(B)} \tag{3.5}$$

where $P(AB)$ is the probability of the product of two events (occurrence of both events A and B) and $P_B(A)$ is called the *conditional probability* of event A, provided that event B has taken place.

For independent events A and B,

$$P_B(A) = P(A) \quad \text{and} \quad P_A(B) = P(B) \tag{3.6}$$

so that

$$P(AB) = P(A)\,P(B) \tag{3.7}$$

Assume that event A can occur together with one and only one of the n mutually exclusive events B_i ($i = 1, 2, \ldots, n$) with the known probability $P(B_i)$. The formula for total probability is

$$P(A) = \sum_{i=1}^{n} P(B_i)\, P_{B_i}(A) \tag{3.8}$$

The events B_i are often called hypotheses of event A and at least one of them must occur for event A to take place. The probability of occurrence of individual hypotheses if event A has occurred, is given by Bayes' formula:

$$P_A(B_i) = \frac{P(B_i A)}{P(A)} = \frac{P(B_i)\, P_{B_i}(A)}{P(A)} = \tag{3.9}$$

$$= \frac{P(B_i)\, P_{B_i}(A)}{\sum_{j=1}^{n} P(B_j)\, P_{B_j}(A)}$$

where $P_A(B_i)$ is the probability of occurrence of the i-th hypothesis together with the occurrence of event A.

3.1.2 Theoretical Probability Distributions of Random Variables

3.1.2.1 Random Variables

The result of a random event is quantitatively described by the value of a *random variable*. The exact value of the random variable cannot be ascertained; only the probability of occurrence of its values can be determined. Random variables are classified as follows:

– *discrete* variables that have, in the given interval, a finite, discrete number of different values,

– *continuous* variables – the set of their values is infinite and continually fills up part of the axis of real numbers.

The *probability distribution* of a random variable describes the allocation of some probability of occurrence to each value of a discrete random variable, or some probability of occurrence of a continuous random variable in a certain interval of possible values.

The probability distribution of a random variable can be described by the *probability density* function (p.d.f.) or by the cumulative distribution function (c.d.f.).

The p.d.f. (Fig. 3.1.a) of a continuous random variable is expressed by function $f(x)$, which is the first derivative of the c.d.f. of the random variable $F(x)$ with respect to x.

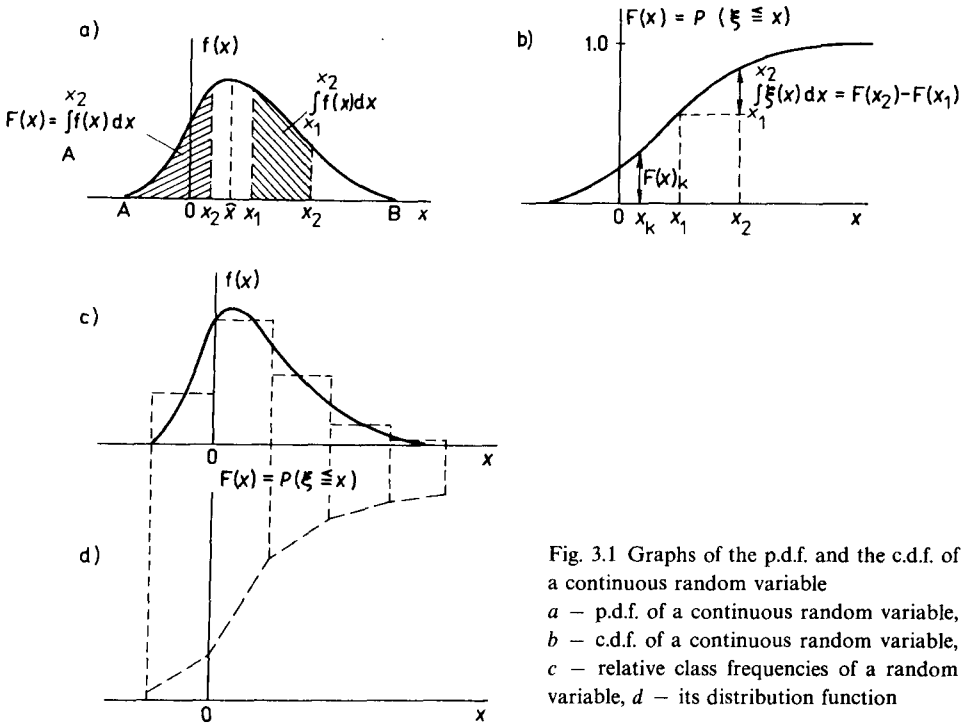


Fig. 3.1 Graphs of the p.d.f. and the c.d.f. of a continuous random variable
a – p.d.f. of a continuous random variable, *b* – c.d.f. of a continuous random variable, *c* – relative class frequencies of a random variable, *d* – its distribution function

The value of the integral

$$\int_{x_1}^{x_2} f(x)\,dx = P(x_1 < \xi \leqq x_2) \tag{3.10}$$

gives the probability of occurrence of the values of the random variable ξ in the interval $\langle x_1; x_2 \rangle$.

The c.d.f., $F(x)$, of a random variable ξ gives the probability that the value of the random variable ξ will not exceed the selected value x, i.e.

$$F(x) = P(\xi \leqq x) \tag{3.11}$$

The c.d.f. (Fig. 3.1.b) of a continuous random variable is continuous, that of a discrete random variable is continuous with the exception of a finite number of points. The c.d.f. of a continuous random variable is a continuous, monotonically increasing function (represented graphically by a curve)

$$F(x) = \int_{-\infty}^{x} f(x)\,dx \tag{3.12}$$

The c.d.f. of a discrete variable is a step function that changes its values only at a finite (or countably infinite) number of points that are the possible values of the random variable ξ and where the jumps occur.

If the sample of discrete random variable values is very large, it is helpful to group the values into classes of equal size. The centre of the class (viz. class mark) represents all the values of the class. It is associated with the class frequency, i.e. the frequency of occurrence of the discrete random variable in the given interval. For each class frequency, n_i, the relative class frequency n_i/n is calculated. The p.d.f. is a step function and the corresponding c.d.f., which is the cumulative function of the relative class frequencies, has the form of a broken line.

The general property of the p.d.f. is

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1 \tag{3.13}$$

3.1.2.2 Numerical Characteristics of Random Variables

Assume a random sample of a certain population of a discrete random variable and a continuous one. There are many problems in mathematical statistics in which it is difficult, or at least not feasible, to determine completely the c.d.f. of a random variable. In such cases it is often possible to describe the distribution of a random variable partially by its moments (i.e. moments of probability distribution). They are classified as a) general (or merely moments) and b) central moments.

a) Moments (*general moments*) are defined by

$$m_k(\xi) = \int_{-\infty}^{\infty} x^k f(x)\,\mathrm{d}x \tag{3.14}$$

where x is the value of the random variable, $f(x)$ is its p.d.f., and k is the order of the moment (k is an integer; this moment is often called the k-th moment of a continuous random variable); or by

$$m_k(\xi) = \sum_{i=1}^{n} x_i^k p_i \tag{3.15}$$

which is the k-th moment of a discrete random variable, where P_i are the probabilities allocated to the values x_i;

b) *Central moments* about the mean μ are defined as the mean (expected) value of $(x - \mu)^k$ (in a mechanical interpretation, the moment is related to the axis which includes the centre of gravity of the graph of the p.d.f.) of a continuous random variable

$$M_k(\xi) = m_k(\xi - \mu) = \int_{-\infty}^{\infty} (x - \mu)^k f(x)\,\mathrm{d}x \tag{3.16}$$

Table 3.1 The main moments of the random variables

Type of moment	Order of moment	Expression for the continuous random variable	Expression for the discrete random variable		Meaning of the moment
			for $p_1 \neq p_2 \neq \ldots \neq p_n$	for $p_1 = p_2 = \ldots = \text{const}$	
General	I.	$m_1(\xi) = \int_{-\infty}^{\infty} x\, f(x)\, dx$	$m_1(\xi) = \sum_{i=1}^{n} x_i p_i$	$m_1 = \frac{1}{n} \sum_{i=1}^{n} x_i$	mean $\mu(x)$ $[\bar{x}]$
Central	I.	$M_1(\xi) = \int_{-\infty}^{\infty} (x - \mu)\, f(x)\, dx$	$M_1(\xi) = \sum_{i=1}^{n} (x_i - \mu)\, p_i$	$M_1(\xi) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})$	–
Central	II.	$M_2(\xi) = \int_{-\infty}^{\infty} (x - \mu)^2\, f(x)\, dx$	$M_2(\xi) = \sum_{i=1}^{n} (x_i - \mu)^2\, p_i$	$M_2(\xi) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$	variance $\sigma^2(x)$ $[s_x^2]$
			$s = \sqrt{M_2(\xi)}$	$s = \sqrt{M_2(\xi)}$	standard deviation $\sigma(x)$ $[s_x]$
				$C_v = \frac{s}{\bar{x}}$	coefficient of variation C_v $[C_{vx}]$
Central	III.	$M_3(\xi) = \int_{-\infty}^{\infty} (x - \mu)^3\, f(x)\, dx$	$M_3(\xi) = \sum_{i=1}^{n} (x_i - \mu)^3\, p_i$	$M_3(\xi) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^3$	–
				$C_s = \frac{M_3(\xi)}{\sqrt{M_2^3(\xi)}}$	coefficient of skewness $C_s(x)$ $[C_{sx}]$
Central	IV.	$M_4(\xi) = \int_{-\infty}^{\infty} (x - \mu)^4\, f(x)\, dx$	$M_4(\xi) = \sum_{i=1}^{n} (x_i - \mu)^4\, p_i$	$M_4(\xi) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^4$	–
				$\gamma = \frac{M_4(\xi)}{M_2^2(\xi)} - 3$	coefficient of kurtosis $E(x)$ $[\gamma]$

or of a discrete random variable

$$M_k(\xi) = \sum_{i=1}^{n} (x_i - \mu)^k p_i \tag{3.17}$$

The meaning of the first moment and the first to the fourth central moments with the expressions for continuous and discrete random variables and for $p = \text{const}$ are given in Table 3.1.

3.1.2.3 Application of Theoretical Probability Distributions

Theoretical probability distributions are used for the evaluation of the type of probability distributions of random events and the values that describe them, and for fitting the empirical probability distributions corresponding to random samples

Table 3.2 Main theoretical probability distributions

Group	Type	Modification (alternative)
Normal		
Normalizing	log-normal	Gauss-Gibrat Galton Ven-Te-Chow
Gamma-distribution		
Beta-distribution		
Pearson	Type I Type III	
Extremal	exponential Weibull Gumbel	
Discrete	geometrical hypergeometrical binomial Poisson	
Involved in analysis of variance	Student – t distribution – χ^2 Snedecor – F	 Fisher

of the population (see 3.1.4). In WRS, *exceedance probability functions* (e.g. flow duration curves) are often used and they are related to the c.d.f. by the formula

$$P(x) = 1 - F(x) = 1 - \int_{-\infty}^{x} f(x)\,\mathrm{d}x \tag{3.18}$$

The empirical frequency function cannot reliably determine the probability of extreme values, even if relatively long time series of observed data are available. Fitting them to the theoretical c.d.f. facilitates extrapolation to the extreme values. The quality and reliability of the results depend on the goodness-of-fit of the selected theoretical c.d.f. with the unknown actual distribution.

Table 3.2. contains the main theoretical c.d.f. used in water management and related tasks.

In fitting the data to the theoretical c.d.f. the following methods are used:

1) By analysis of the characteristics and properties of the random variable and its origin (analysis of the boundary conditions, skewness of the empirical c.d.f., etc.) a feasible type, or several types of theoretical c.d.f. are determined.

2) By means of the parameters of the empirical c.d.f. (sample values) the parameters of the theoretical c.d.f. are estimated.

3) The goodness-of-fit between the empirical and theoretical c.d.f.s is evaluated.

3.1.2.4 The Normal Probability Distribution

The normal (Gaussian) probability distribution is the most important one in probability theory and in mathematical statistics, as it is suitable for the approximation of empirical p.d.f. of quantities observed in different phenomena, both continuous and discrete, and also for the approximation of other, more complicated, theoretical probability distributions. A random variable, the values of which are the sum of a large number of mutually independent effects, each having a small impact, has the normal probability distribution.

A further condition for normal distribution is a large number of elements in a sample. It is necessary to differentiate between a large number of elements in a sample and a large number of independent effects. In general, it can be said that a large number of elements in the sample contributes to the goodness-of-fit between the empirical and theoretical distributions, whereas the number of independent effects determines the type of theoretical distribution adequate for the given random variable.

The normal distribution of a continuous random variable has two parameters: the mean μ and the standard deviation σ. It can also be used for a discrete random variable.

The p.d.f. and c.d.f. of the normal distribution are presented in general and standardized forms in Table 3.3.

Table 3.3 Normal probability distribution

random variable	x
mean	$\mu(x)$
standard deviation	$\sigma(x)$
p.d.f.	$f(x) = c\, e^{-[x-\mu(x)]^2/2\sigma^2(x)}$; $c = \dfrac{1}{\sigma(x)\sqrt{2\pi}}$
c.d.f.	$F(x) = \dfrac{1}{\sigma(x)\sqrt{2\pi}} \int_{-\infty}^{x} e^{-[x-\mu(x)]^2/2\sigma^2(x)}\, dx$
standardized variable	$t = \dfrac{x - \mu(x)}{\sigma(x)}$
p.d.f. of the standardized variable	$f(t) = \dfrac{1}{\sqrt{2\pi}} e^{-t^2/2}$
c.d.f. of the standardized variable	$F(t) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-t^2/2}\, dt$

It is convenient to refer to the normal distribution with the general form of p.d.f. (Table 3.3) as the distribution $N(\mu, \sigma^2)$, and the standardized form of p.d.f. as $N(0, 1)$.

The graph of the p.d.f. and c.d.f. for some selected values of σ and $\mu = 0$ is given in Fig. 3.2.

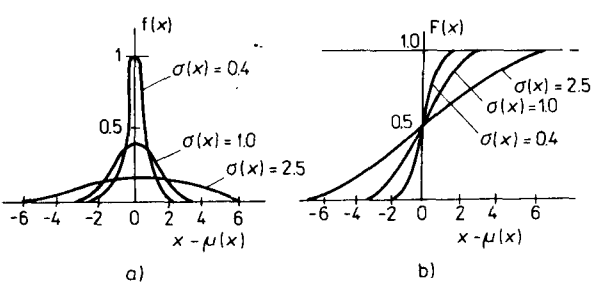


Fig. 3.2 Normal probability distribution of the random variable for different values of $\sigma(x)$
a – p.d.f., *b* – c.d.f.

In WRS contexts the probability distributions are frequently not symmetrical and they are bounded on one or both sides. The normal distribution, however, is an approximative basis used for the derivation of other convenient distributions and for the development of methods of probability and the statistical treatment of random samples.

Table I.1 Class frequencies

i	Class interval	x_i'	n_i
1	8.88 – 9.12	9.00	1
2	9.13 – 9.37	9.25	3
3	9.38 – 9.62	9.50	11
4	9.63 – 9.87	9.75	24
5	9.88 – 10.12	10.00	42
6	10.13 – 10.37	10.25	62
7	10.38 – 10.62	10.50	58
8	10.63 – 10.87	10.75	44
9	10.88 – 11.12	11.00	19
10	11.13 – 11.37	11.25	9
11	11.38 – 11.62	11.50	4
12	11.63 – 11.87	11.75	2
13	11.88 – 12.12	12.00	1
			280

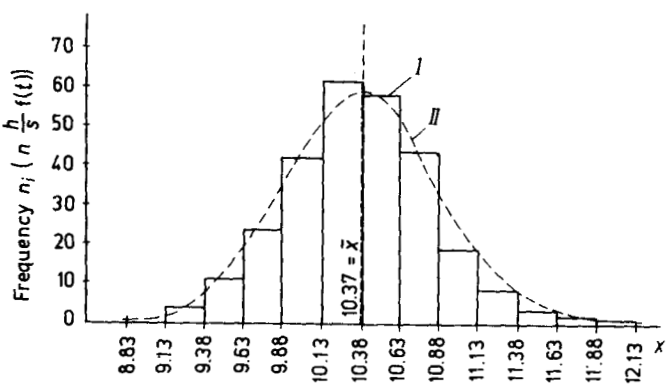


Fig. I.1 Fitting of the theoretical p.d.f. (II) to the histogram of class frequencies (I)
I – histogram of class frequencies, *II* – theoretical normal p.d.f.

Example I

A sample of 280 elements is given by its class frequencies (Table I.1) with the sample mean $\bar{x} = 10.37$ and the sample standard deviation $s = 0.475$. A theoretical normal p.d.f. is to be fitted to it.

Solution

The empirical p.d.f. is clearly close to the normal p.d.f. The values of the theoretical normal p.d.f. are calculated in Table I.2. The values $t = (x_i - \bar{x})/s$ are the values of the standardized variable where x_i are the lower class limits for which the theoretical values of the p.d.f. are calculated.

Table I.2 Computation of the theoretical normal p.d.f.

i	x_i	$x_i - \bar{x}$	$t = \frac{x_i - \bar{x}}{s}$	$f(t)$	$\frac{nh}{s} f(t)$
1	8.88	−1.493 2	−3.14	0.002 9	0.426
2	9.13	−1.243 2	−2.61	0.013 3	1.95
3	9.38	−0.993 2	−2.08	0.046 0	6.77
4	9.63	−0.743 2	−1.55	0.120 2	17.5
5	9.88	−0.493 2	−1.03	0.234 8	34.5
6	10.13	−0.243 2	−0.51	0.350 2	51.5
7	10.38	+0.006 8	+0.014	0.394 0	58.0
8	10.63	+0.256 8	+0.54	0.344 5	50.6
9	10.88	+0.506 8	+1.06	0.227 5	33.5
10	11.13	+0.756 8	+1.59	0.112 8	18.2
11	11.38	+1.006 8	+2.12	0.041 5	6.10
12	11.63	+1.256 8	+2.64	0.012 2	1.79
13	11.88	+1.506 8	+3.17	0.002 7	0.397

The values of $f(t)$ corresponding to values t are obtained from tables (e.g. Reisenauer, 1970). These values are multiplied by the value of the following expression

$$\frac{nh}{s} = \frac{280 \cdot 0.25}{0.475}$$

where h is the class length.

The values of the class frequencies and theoretical p.d.f. are given in Table I.1.

3.1.2.5 Normalizing Probability Distributions

The normal probability distribution has a simple mathematical representation and tables on the p.d.f. and c.d.f. of standardized variables are available (the computer programs for their calculation are included in the software of each modern computer). Therefore, there have been many attempts to transform skew probability distributions of a random variable x to a variable y that is normally distributed.

The principle of this transformation is the determination of a function φ for *normalizing the transformation* $y = \varphi(x)$.

The simplest transformation function is $y = x^a$. In hydrology, C. K. Stidd used the exponent $a = \frac{1}{3}$, which proved to be helpful for various types of time series of precipitation (e.g. annual, monthly and daily precipitation) (Kendall, 1970).

The *logarithmic transformation* expressed by Johnson in a general form is often used:

$$y = a + b\left(\log \frac{x - m}{l}\right) \tag{3.19}$$

This formula is the most general and includes other formulae (e.g. Galton's) as a special case. An apparent disadvantage is its large number of parameters (a, b, l, m), but on the other hand it has the advantage of good adaptability to a wide range of skew distributions.

Modifications of the log-normal distribution applied in meteorology, hydrology and water management have been developed by F. Galton, R. Gibrat and Ven Te Chow.

The log-normal probability distribution of a random variable x is such that the variable, after the logarithmic transformation (e.g. (3.19)), has the normal probability distribution.

Table 3.4 Log-normal probability distribution

p.d.f.	$f(x) = \dfrac{1}{(x - x_0)\,\sigma(y)\sqrt{2\pi}} \exp\{-[\lg(x - x_0) - \mu(y)]^2/2\sigma^2(y)\}$
form of logarithmic transformation (Galton)	$y = \lg\lvert x - x_0\rvert$
mean	$\mu(y) = \lg\sigma(x) - \lg\lvert C_\gamma\rvert - \frac{1}{2}\lg(1 + C_\gamma^2)$
variance	$\sigma^2(y) = \lg(1 + C_\gamma^2)$
coefficient of skewness	$C_s(x) = C_\gamma^3 + 3C_\gamma$

The log-normal distribution proved to be adequate for many skew distributions that correspond, for example, to maximum flows (frequently $C_s > 3C_v$, i.e. the coefficient of skewness is greater than three coefficients of variation) and to monthly flows.

The analytical expression of the p.d.f. and relationships describing the basic statistical characteristics are given in Table 3.4 for Galton's transformation of the random variable

$$y = \lg |x - x_0| \tag{3.20}$$

Further transformations were used by R. Gibrat (Gauss–Gibrat distribution)

$$y = a \log (x - x_0) + b \tag{3.21}$$

Ven Te Chow

$$y = \ln x \tag{3.22}$$

and F. Galton (a new formula)

$$y = \alpha(\lg x - \beta) \tag{3.23}$$

The Gauss–Gibrat distribution fitted the probability distributions of daily flows and annual maximum flows well (Votruba and Broža, 1966). Gumbel's extreme probability distribution is a special case of the log-normal probability distribution (Ven Te Chow, 1964).

3.1.2.6 Gamma Probability Distribution

The mathematical expression of the p.d.f. of the gamma distribution uses the gamma function, i.e. Euler's function of second order $\Gamma(x)$. It is defined by the integral

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} \, dt \tag{3.24}$$

that is convergent for $x > 0$. Its form can be described by the expressions

$$\lim_{x \to \infty} \Gamma(x) = +\infty \quad \text{and} \quad \lim_{x \to 0+} \Gamma(x) = +\infty$$

and it has a local minimum for $x \simeq 1.46$.

Further properties of the gamma function are

a) $\Gamma(x + 1) = x\,\Gamma(x)$
b) $\Gamma(n) = (n - 1)!$ for positive integer n (3.25)

The gamma function has been tabulated (e.g. Abramowitz *et al.*, 1964; Rektorys *et al.*, 1963), and it is used for the calculation of many difficult integrals that can be transformed to it.

Pearson and Weibull used the gamma function for the expression of their distri-

butions. The gamma distribution (or more precisely the incomplete gamma distribution) was used by G. L. Barger and H. C. S. Thom in processing precipitation time series (Kendal, 1970). Due to the possibility of selections many combinations of its two parameters with a pronounced effect on the shape of the p.d.f. (Fig. 3.3), further applications in water management are envisaged.

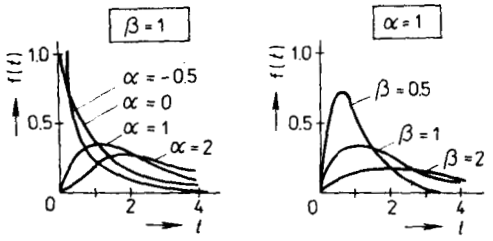


Fig. 3.3 P.d.f. of the gamma-distribution for different combinations of α and β

The general mathematical form of the p.d.f. of the random variable x, i.e. mean, standard deviation and mode, is given in Table 3.5. This distribution has been tabulated by K. Pearson, and it is published not only in specialist mathematical literature but also in water management references. For the estimation of parameters the method of maximum likelihood is recommended and preferred to the method of moments.

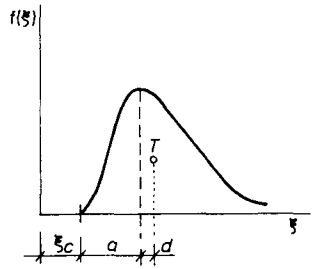
Table 3.5 Gamma probability distribution

p.d.f.	$f(x) = \begin{cases} \dfrac{1}{\alpha!\beta^{\alpha+1}} x^{\alpha} e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{for } x \leqq 0 \end{cases}$
c.d.f.	$F(x) = \dfrac{1}{\alpha!} \Gamma_{x/\beta}(\alpha + 1)$
mean	$\mu(x) = \beta(\alpha + 1)$
standard deviation	$\sigma(x) = \beta\sqrt{\alpha + 1}$
coefficient of skewness	$C_s(x) = \dfrac{2}{\sqrt{\alpha + 1}}$

α, β – parameters of the gamma probability distribution in a general form

$\Gamma_{x/\beta}(\alpha + 1)$ – gamma-function

Table 3.6 The Pearson probability distribution – type III

Quantity		General form	Standardized form	Geometrical interpretation
random variable		x	$t = \dfrac{x - \mu(x)}{\sigma(x)}$	ξ
mean		$\mu(x)$	0	$\mu(\xi) = \xi_0 + a + d$
standard deviation		$\sigma(x)$	1	$\sigma(\xi) = \sqrt{d(a+d)}$
coefficient of skewness		$C_s = \dfrac{2}{\alpha}$	$C_s = \dfrac{2}{\alpha}$	$C_s = 2\sqrt{\dfrac{d}{a+d}}$
p.d.f.		$f(x) = \dfrac{\alpha^{\alpha^2} \exp\left(\dfrac{\alpha\mu(x)}{\sigma(x)} - \alpha^2\right)}{\Gamma(\alpha^2)\,\sigma^{\alpha^2}(x)} \cdot$ $\cdot [\alpha\sigma(x) - \mu(x) + x]^{\alpha^2-1} \, e^{-(\alpha/\sigma(x))x}$	$f(t) = \dfrac{\alpha^{\alpha^2} e^{-\alpha^2}}{\Gamma(\alpha^2)} \cdot$ $\cdot (\alpha + t)^{\alpha^2-1} e^{-\alpha t}$	$f(\xi) = \dfrac{1}{d(a/d+1)\,\Gamma(a/d+1)} \cdot$ $\cdot (\xi - \xi_0)^{a/d} e^{(-\xi-\xi_0)/d}$
the range of the random variable	$\alpha > 0$	$[\mu(x) - \alpha\sigma(x); +\infty]$	$(-\alpha; \infty)$	
	$\alpha < 0$	$[-\infty; \mu(x) - \alpha\sigma(x)]$	$(-\infty; -\alpha)$	
schematic geometrical representation of the p.d.f.				

3.1.2.7 The Pearson Probability Distribution

K. Pearson derived the p.d.f. on the basis of a general differential equation

$$\frac{dy}{dx} = \frac{x + a}{b_0 + b_1 x + b_2 x^2} y \tag{3.26}$$

with coefficients a, b_0, b_1, b_2, which are real numbers. According to the values of these coefficients and their combinations, he described 12 types of theoretical distributions, derived, in principle, from the general beta distribution. These functions have a purely theoretical basis; however, in many cases they can be successfully used to fit empirical distributions. Sometimes they are considered a good universal type of skew distributions, adequate for discrete random variables (e.g. approximation of the binomial distribution).

The general form of the p.d.f. of the Pearson distribution is given by the expression

$$f(x) = \exp\left(\int_{-\infty}^{x} \frac{x + a}{b_0 + b_1 + b_2 x^2} \, dx\right) \tag{3.27}$$

The Pearson probability distribution type III is a type of skew distribution with a lower bounded, bell-shaped form. From the analytical description of the p.d.f. in Table 3.6 it is clear that it has three parameters: the mean $\mu(x)$, standard deviation $\sigma(x)$, and the constant α, which defines the coefficient of skewness. The normal distribution is a special case of the Pearson probability distribution type III with a zero skewness.

3.1.2.8 Exponential Probability Distribution

Exponential probability distribution is a single-parameter probability distribution, and it is given in Table 3.7 by its p.d.f. and c.d.f. It can be considered a special case of the gamma distribution, possibly also of the Weibull distribution.

This function is widely used in reliability theory. However, in hydrology and in water management it has been successfuly applied in the statistical treatment of the maximum flows in a given period (Votruba and Broža, 1966).

For application to water management the c.d.f. can be written in a general form

$$F(z) = 1 - e^{-z} \tag{3.28}$$

where

$$z = \varphi(x) \tag{3.29}$$

Goodrich (Votruba and Nacházel, 1969) used the transformation of the type

$$z = k(x - x')^{1/\alpha} \tag{3.30}$$

with three parameters (constants) k, x', α.

Table 3.7 Exponential probability distribution

p.d.f.	$f(x) = \begin{cases} \Theta^{-1} e^{-x/\Theta} & \text{for } x \geqq 0 \\ 0 & \text{for } x < 0 \end{cases}$
c.d.f.	$F(x) = \begin{cases} 1 - e^{-x/\Theta} & \text{for } x \geqq 0 \\ 0 & \text{for } x < 0 \end{cases}$
mean	$\mu(x) = \Theta$; Θ – parameter of distribution
standard deviation	$\sigma(x) = \Theta$
coefficient of skewness	$C_s(x) = 2$

3.1.2.9 The Gumbel Probability Distribution

The Gumbel probability distribution is a type of extreme probability distribution, which is suitable for the maximum values of samples. In hydrology and in WRS problems it has been used for processing flood flows (Coutagne, 1951).

Table 3.8 The Gumbel probability distribution

p.d.f.	$f(x) = \exp(-z) \exp[-\exp(-z)]$
c.d.f.	$F(x) = \exp[-\exp(-z)]$
equation of the exceedance probability curve	$P(x) = 1 - \exp[-\exp(-z)]$
transformation of random variable	$z = \alpha(x - \beta)$
constants in transformation	$\alpha = \dfrac{\pi}{\dot{\sigma}(x)\sqrt{6}}$ $\beta = \dot{\mu}(x) - \dfrac{\gamma}{\alpha}$

α, β – constants in transformation
$\pi = 3.14 \ldots$ – Ludolf's number
$\gamma = 0.577 \ldots$ – Euler's number
$\dot{\mu}(x), \dot{\sigma}(x)$ – point estimates of the mean and standard deviation of the population

Table 3.9 The method of fitting theoretical exceedance probability functions by quantiles method

Probability distribution	Quantiles estimated from empirical probability curve	Auxiliary parameters	Estimation of parameters of the exceedance probability curve	Equation of the theoretical exceedance probability function
log-normal	x_5, x_{50}, x_{95}	$x_0 = \dfrac{x_5 x_{95} - x_{50}^2}{x_5 + x_{95} - 2x_{50}}$	$s_y = 0.304 \log \dfrac{x_5 - x_0}{x_{95} - x_0}$	$\log(x_p - x_0) = \log(x_{50} - x_0) + s_y \Phi_B(P, C_s)$ for $C_s = 0$
binomial	x_5, x_{50}, x_{95}	$S = \dfrac{x_5 + x_{95} - 2x_{50}}{x_5 - x_{95}} = \dfrac{\Phi_5 + \Phi_{95} - 2\Phi_{50}}{\Phi_5 - \Phi_{95}}$	$\bar{x} = x_{50} - s_x \cdot \Phi_{50}$ $s_x = \dfrac{x_5 - x_{95}}{\Phi_5 - \Phi_{95}}$; $C_v = \dfrac{s_x}{\bar{x}}$	$x_p = \bar{x}[1 + C_s \Phi_B(P, C_s)]$
Gumbel	x_5, x_{95}	$\alpha = \dfrac{4.067}{x_5 + x_{95}}$ $\beta = 0.27x_s + 0.73x_{95}$	$\bar{x} = 0.412x_5 + 0.5x_{95}$ $s_x = 0.315(x_5 - x_{95})$	$x_p = \beta + \dfrac{1}{\alpha} Z_p = \bar{x} + s_x \Phi_G(P, C_s)$
exponential	x_5, x_{50}, x_{95}	$S_z = \dfrac{2 \log x_{50} - \log x_5 - \log x_{95}}{\log x_5 - \log x_{95}}$ a) for $C_{sz} > 0$: $\alpha = \dfrac{\log k_5 - \log k_{95}}{\log x_5 - \log x_{95}}$ b) for $C_{sz} < 0$: $\alpha' = \dfrac{\log k_{95} - \log k_5}{\log x_5 - \log x_{95}}$ k_5, k_{95} tabulated for C_{sz}	$\log \bar{z} = \alpha \log x_{50} - \log k_{50}$ $\log \bar{z} = \alpha' \log x_{50} - \log k_{50}$	$\log z_p = \log \bar{z} + \log[1 + 0.5 C_{sz} \Phi_B(P, C_{sz})]$ $\log x_p = \dfrac{1}{\alpha} \log z_p$

Under Czechoslovak conditions this distribution was applied in processing maximum precipitation (for n-day, n-hour values). In view of the analytical form of the p.d.f. (Table 3.8) it is also called double-exponential probability distribution, with the standardized deviation about the mode as the argument. E. J. Gumbel called it the law of maximum values as it was derived from an investigation of the highest age. It is skew (with one constant value of the skewness coefficient $C_s = 1.139$) and unbounded.

3.1.2.10 Estimation of Parameters of Cumulative Distribution Function

Analyses performed in WRS studies have shown that for the determination of the parameters of the c.d.f. (or exceedance probability curves) the method of moments need not necessarily be the best one. Therefore some other methods have been developed for a simpler, less cumbersome and often more reliable estimation of the parameters of the c.d.f. (Alexeyev, 1960). These methods use *quantiles*, i.e. values

Table 3.10 The used probability papers

Type of network	Scale of the abscissa (probability)	Scale of the ordinate
normal distribution	derived from the normal c.d.f. (symmetrically increasing on both sides)	linear
log-normal	derived from the c.d.f.	logarithmic
Gumbel	derived from the straight form of the relation between the standardized variable and the c.d.f. (skew, it increases to zero more than to one)	linear
Fréchet	-"-	logarithmic
Goodrich (exponential distribution)	log log $(1/P)$ (P ... derived from the c.d.f. of the normal distribution)	logarithmic
Brovkovich	derived from the c.d.f. of normal distribution (symmetrically increasing on both sides)	derived from the straight form of the standardized variable and the c.d.f. of binomial distribution with $C_s = 2C_v$

chosen in such a way that the observation less than this value forms a given (required or chosen) part of the sample (more correctly: for the number p the p-th quantile x_p of the continuous random variable x having c.d.f. $F(x)$ is defined by the smallest number for which $F(x_p) = p$). The method of quantiles is used for log-normal, binomial, Gumbel and exponential distributions. It is often based on values of two or three quantiles estimated from the smoothed empirical c.d.f. (exceedance probability function) and standardized values of the c.d.f. $\Phi\ (P, C_s)$ which have been tabulated for different types of probability distributions.

The equation for determining the basic statistical characteristics and the value of the theoretical exceedance probability function determined by the quantiles method and the corresponding standardized variables are given in Table 3.9.

The fitting of the theoretical c.d.f. (exceedance probability function) to the empirical one in hydrology is often performed in the following way: the values of the random variable are ranked in the ascending (descending) order, the empirical probabilities (probabilities of exceedance) are calculated by the plotting – position – probability formula $p = (m + a)/(n + b)$ where m is the rank, n is the number of ranked events, a and b are constants (e.g. Chegodayev has $a = -0.3$, $b = 0.4$, Hazen, $a = -0.5$, $b = 0$ or Weibull, $a = 0$, $b = 1$).

These values p and the corresponding values of stochastic variables are plotted on probability paper, where the scale of probability typically increases towards both sides. The exceedance probability function (or c.d.f.) is represented on this paper by a curve approaching a straight line, or by a straight line if the sample investigated has the theoretical c.d.f. used in the construction of the probability network. Therefore, the probability paper that seems to correspond best to the empirical sample is selected. Table 3.10 shows the main probability papers used in WRS with a description of the scales on the axes.

Example II

Earth samples were taken during the construction of an earthen dam to provide a sample of values for the density of wet earth. The frequencies (histogram) in Fig. II.1 show a negative skewness that suggests the use of binomial distribution. The parameters of the fitted theoretical c.d.f. will be determined using the method of quantiles.

Solution

The values of the quantiles are estimated from the empirical exceedance probability curve (Fig. II.2)

$$x_5 = 2.190 \quad x_{50} = 2.080 \quad x_{95} = 1.937$$

An auxiliary value S is calculated

$$S = \frac{x_5 + x_{95} - 2x_{50}}{x_5 - x_{95}} = \frac{2.190 + 1.937 - 2 \cdot 2.080}{2.190 - 1.937} = -0.130 \qquad \text{(II.1)}$$

From the tables (e.g. Votruba *et al.*, 1973, appendix II) we get $S = -0.13$

$$C_s = -0.470; \quad \Phi_5 - \Phi_{95} = 3.260; \quad \Phi_{50} = -0.080$$

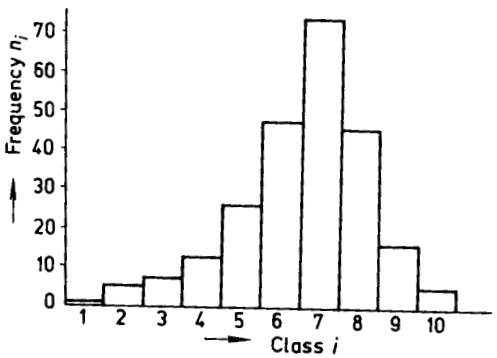


Fig. II.1 Histogram of class frequencies of earth density in a wet stage

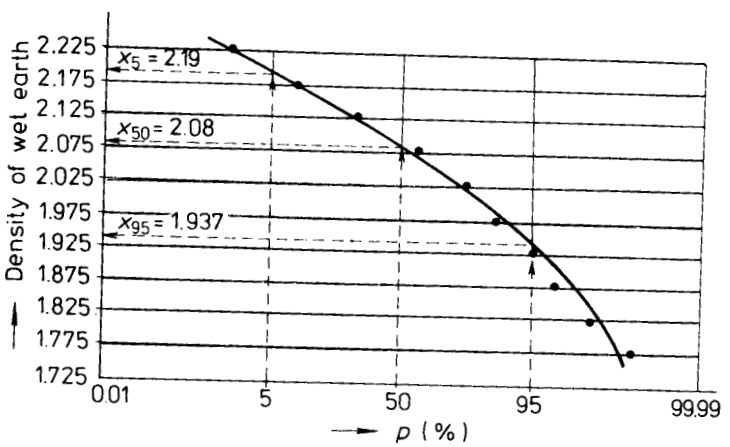


Fig. II.2 Smoothed empirical exceedance probability curve

Table II.1 The class frequencies of the density of wet earth

i	Value of the class mark $[x_i]$	Frequency $[n_i]$	Rank of the last element in each class $[m]$	Exceedance probability $p = \frac{m - 0.3}{n + 0.4}\ 100\%$
1	1.775	1	239	99.8
2	1.825	4	238	99.2
3	1.875	7	234	97.3
4	1.925	13	227	94.6
5	1.975	26	214	89.1
6	2.025	48	188	78.4
7	2.075	74	140	58.3
8	2.125	46	66	27.4
9	2.175	16	20	8.22
10	2.225	4	4	1.54

The sample parameters of the exceedance probability function are calculated and used as the estimators of parameters of the theoretical probability distribution (Table II.1)

$$s_x = \frac{x_5 - x_{95}}{\Phi_5 - \Phi_{95}} = \frac{2.190 - 1.937}{3.260} = 0.078 \tag{II.2}$$

$$\bar{x} = x_{50} - s_x \Phi_{50} = 2.080 - 0.078(-0.08) = 2.086 \tag{II.3}$$

$$C_v = \frac{s_x}{\bar{x}} = \frac{0.078}{2.086} = 0.038 \tag{II.4}$$

3.1.3 Dependence and Correlation

3.1.3.1 Functional and Probability Relationships

Among the variables (including random ones) describing the processes in the natural and technical sciences there exist functional and probability (stochastic) relationships; in water management a linear correlation is the most frequent one.

The investigation of the probability relationships of the random variables was facilitated by the following factors: the accumulation of sufficient information con-

cerning the phenomena and describing random variables; a deeper understanding of the genesis of the processes investigated, which helped to reduce possible errors in the conclusions concerning the existence and nature of the probability relationships, improvement of methods of analysis; and, last but not least, the use of computers, for their evaluation.

Assume that the phenomena investigated have, in the simplest case, a single property that can be measured and expressed quantitatively. The quantity that describes this property is, in general, variable in time, t, and in space, s, and includes some random component ε, so that

$$x = f(t) + t_0 + \varepsilon_1 \tag{3.31}$$

and, at the same time,

$$x = g(s) + s_0 + \varepsilon_2 \tag{3.32}$$

or

$$x = h(s, t) + s_0 + t_0 + \varepsilon_3 \tag{3.33}$$

where s_0 and t_0 are constants

Similarly, a variable Y can be expressed. Assume that the variables X and Y are function of more random variables $Z_1, Z_2, \ldots, Z_n$, which are common to both variables, and that Y is also a function of $U_1, U_2, \ldots, U_r$ and Y a function of $V_1, V_2, \ldots, V_s$; then there is a probability relationship between the variables X and Y.

In WRS contexts, probability relationships between the variables investigated are quite common. If overlooked, this could lead to incorrect results with serious technical and economic consequences.

3.1.3.2 Multivariate Probability Distribution

In section 3.1.2, one-dimensional probability distributions were investigated where the probability of occurrence of the random variable can be described by a single value. In the investigation of probability relationships among random variables it is useful to start with the idea of multivariate probability distributions which take into account more factors of each element of the statistical sample. The probability of their occurrence is not given by a single value but by several values depending on the occurrence of a certain factor; in fact, this probability is a conditional probability in the sense of section 3.1.1.

Let us start with the two-variate probability distribution of the random variable (X, Y), where $X = x_i\,(i = 1, 2, \ldots, n)$ and $Y = y_j\,(j = 1, 2, \ldots, m)$, the matrix on which

is given in Table 3.11. Each row of the matrix gives the conditional probability distribution of the random variable X, given the value Y, and each column the conditional probability of the random variable Y, given X. The last row and last column contain the values of the unconditional probability distributions of the values X and Y respectively.

Table 3.11 Schematic representation of the matrix of the probability distribution of a two-variate discrete random variable

Y \ X	x_1	x_2	...	x_i	...	x_n	
y_1	$p(x_1, y_1)$	$p(x_2, y_1)$	...	$p(x_i, y_1)$	...	$p(x_n, y_1)$	$p(y_1)$
y_2	$p(x_1, y_2)$	$p(x_2, y_2)$	...	$p(x_i, y_2)$	...	$p(x_n, y_2)$	$p(y_2)$
⋮							
y_j	$p(x_1, y_j)$	$p(x_2, y_j)$	...	$p(x_i, y_j)$	...	$p(x_n, y_j)$	$p(y_j)$
⋮							
y_m	$p(x_1, y_m)$	$p(x_2, y_m)$	...	$p(x_i, y_m)$	...	$p(x_n, y_m)$	$p(y_m)$
	$p(x_1)$	$p(x_2)$		$p(x_i)$		$p(x_n)$	

The k-th and c-th moments and central moments of the two-variate random variables (X, Y) are determined by the expressions

$$ {}_{x,y}m_{k,c} = \frac{1}{n} \sum_{i=1}^{n} x_i^k y_j^c \tag{3.34} $$

$$ {}_{x,y}M_{k,c} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^k (y_i - \bar{y})^c \tag{3.35} $$

If $k \neq 0$ and at the same time $c \neq 0$, then the moments are called *mixed product moments*; the most important of these is the first central moment ${}_{x,y}M_{1,1}$ called the *covariance*

$$ {}_{x,y}M_{1,1} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \tag{3.36} $$

All the covariances of a n-dimensional sample of the random variables $X_1, X_2, \ldots, X_n$ can be arranged in a square matrix of the following form:

$$ {}_{i,j}\mathbf{M} = \begin{Vmatrix} {}_{1,1}M & {}_{1,2}M & \cdots & {}_{1,n}M \\ {}_{2,1}M & {}_{2,2}M & \cdots & {}_{2,n}M \\ \cdots & & & \\ {}_{n,1}M & {}_{n,2}M & \cdots & {}_{n,n}M \end{Vmatrix} \tag{3.37} $$

which is called a *covariance matrix*. It is symmetrical with respect to the main diagonal so that it can be written as the upper triangle matrix

$$ {}_{i,j}\mathbf{M} = \begin{Vmatrix} {}_{1,1}M & {}_{1,2}M & \cdots & {}_{1,n}M \\ & {}_{2,2}M & \cdots & {}_{2,n}M \\ & & \ddots & \vdots \\ & & & {}_{n,n}M \end{Vmatrix} \tag{3.38} $$

For the independent (uncorrelated) random variables $X_1, X_2, \ldots, X_n$ this matrix is, in fact, a diagonal matrix with non-zero elements on the main diagonal only:

$$ {}_{i,j}\mathbf{M} = \begin{Vmatrix} {}_{1,1}M & 0 & \cdots & 0 \\ & {}_{2,2}M & \cdots & 0 \\ & & \ddots & \vdots \\ & & & {}_{n,n}M \end{Vmatrix} \tag{3.39} $$

Instead of the covariance matrix, a *correlation matrix* is often used for the standardized variables. Its elements are not the covariances but the *correlation coefficients* defined by

$$ r_{ij} = \frac{{}_{i,j}M_{1,1}}{S_i S_j} \tag{3.40} $$

so that (3.38) becomes (3.41)

$$ \mathbf{r}_{i,j} = \begin{Vmatrix} 1 & r_{1,2} & r_{1,3} & \cdots & r_{1,n} \\ & 1 & r_{2,3} & \cdots & r_{2,n} \\ & & 1 & \cdots & r_{3,n} \\ & & & \cdots & \\ & & & & 1 \end{Vmatrix} \tag{3.41} $$

3.1.3.3 Methods of Determination of the Linear Statistical Dependence Between Random Variables

These methods encompass the calculation of the coefficients of linear regression and correlation from the covariances, the coefficient of rank correlation, and also the estimation of the statistical dependence by graphical representation of the regres-

sion relationship by plotted points. The *coefficient of linear correlation*, r, is a measure of the statistical dependence between the random variables X and Y (their values x_i and y_i) and it is calculated from the covariances by the following equations

$$r = \frac{{}_{x,y}M_{1,1}}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.42}$$

$-1 \leqq r \leqq 1$, and for uncorrelated (i.e. statistically independent) random variables X and Y, $r = 0$.

The statistical dependence can be either *direct* (i.e. positive for $0 < r \leqq 1$) or *indirect* (i.e. negative for $-1 \leqq r < 0$). The value of the correlation coefficient is invariant to linear transformation of random variables.

The coefficient of linear correlation is a good measure of the statistical dependence of random variables X and Y, if the relationship between them approaches the linear one. Therefore it is useful to evaluate the degree of linearity of the relationship between two random variables (e.g. by plotting of points).

The following formulation is recommended for the calculation of the correlation coefficient:

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}} \tag{3.43}$$

The *rank correlation coefficient* (Spearman's correlation coefficient) used for testing the linear dependence of two random variables, is given by the formula

$$r_p = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{3.44}$$

For its calculation, the values x_i and y_i are replaced by their ranks (in descending order). The difference of ranks of the corresponding values v_{x_i} and v_{y_i} in both samples of random variables X and Y are denoted by d_i, n being the number of correlated pairs. To check the correct determination of the values d_i, the condition

$$\sum_{i=1}^{n} d_i = 0$$

may be used. The values of the rank coefficient have the same range and can be compared with the ordinary correlation coefficient determined by covariances.

Example III

Calculate the value of rank correlation coefficient of the mean annual flows Q_a in the period 1931 – 1960 in the town of Písek on the River Otava and compare it with the ordinary correlation coefficient.

Table III.1 Calculation of the rank correlation coefficient

i	Q_r ($m^3 s^{-1}$) a)	v_{x_i}	v_{y_i}	$d_i = v_{x_i} - v_{y_i}$	d_i^2
1	40.2	6	8	− 2	4
2	37.6	8	26	−18	324
3	35.2	26	28	− 2	4
4	31,0	28	22	+ 6	36
5	28.2	22	19	+ 3	9
6	27.5	19	15	+ 4	16
7	27.1	15	10	+ 5	25
8	26.4	10	3	+ 7	49
9	25.6	3	1	+ 2	4
10	24.9	1	2	− 1	1
11	23.8	2	14	−12	144
12	23.4	14	27	−13	169
13	22.7	27	5	+22	484
14	21.8 (2×)	5	11	− 6	36
15	21.5	11	12	− 1	1
16	21.1	12	23	−11	121
17	20.5	23	9	+14	196
18	20.3	9	16	− 7	49
19	19.7	16	27	−11	121
20	19.4	27	24	+ 3	9
21	17.8	24	25	− 1	1
22	17.2	25	21	+ 4	16
23	17.1	21	14	+ 7	49
24	16.3	14	4	+10	100
25	16.2	4	17	−13	169
26	14.1	17	13	+ 4	16
27	12.4 (2×)	13	7	+ 6	36
28	10.4	7	18	−11	121
29		18	20	− 2	4
Σ				×)	2314

a) ranked in descending order

×) the sum $\sum d_i$ is not equal to zero as in the column v_{x_i} in the computation of the autocorrelation coefficient the 30th year element is not included, which has $v_{x_{30}} = 20$; similarly in the column v_{y_i} the 1st element is missing, which has $v_{y_1} = 6$. Then $(-\sum d_i = -14) + 20 - 6 = 0$

Solution

The calculation was arranged in a tabular form (Table III.I.). The coefficient of the rank correlation was calculated by formula (3.44)

$$r_p = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 2314}{29(29^2 - 1)} = 0.428 \qquad \text{(III.1.)}$$

The value of the ordinary correlation coefficient using formula (3.42) or (3.43) is a little higher; $r = 0.497$.

The correlation technique is often accompanied by regression analysis. Linear regression functions (geometrically represented by straight lines) relate a dependent variable X to an independent variable Y or vice versa. The plotted points of the relationship are concentrated along a straight line if the absolute value of the correlation coefficient converges to one (Fig. 3.4).

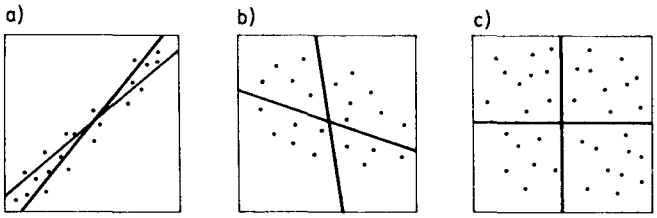


Fig. 3.4 Relative position of the regression straight lines
a – close direct linear regression, *b* – loose indirect linear regression, *c* – independence

The general equation for the linear regression (equation of the regression straight line) of the dependent random variable Y and independent random variable X is

$$Y = a + bX \qquad (3.45)$$

where a (intercept) and b (regression coefficient) are constants that are determined by the least squares method from the sum

$$\sum_{i=1}^{n} y_i = na + b \sum_{i=1}^{n} x_i \qquad (3.46)$$

The following expressions for these constants have been derived

$$a = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2} \qquad (3.47)$$

$$b = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2} \tag{3.48}$$

By substituting these expressions into eq. (3.45), the equation of the linear regression, the expanded function $Y = f(x)$ is obtained. Similarly, the contents a' and b' can be calculated for the function $x = g(Y)$ in the form

$$X = a' + b' Y \tag{3.49}$$

The constant b in equation (3.45) is the *regression coefficient* of the linear function $Y = f(X)$, and b' in equation (3.49) the regression coefficient of the linear function $X = g(Y)$. They can be related to the standard deviations s_x and s_y of the random variables X and Y in the equations

$$b = r \frac{s_y}{s_x} \tag{3.50}$$

$$b' = r \frac{s_x}{s_y} \tag{3.51}$$

and the regression function becomes

$$y - \bar{y} = r \frac{s_y}{s_x} (x - \bar{x}) \tag{3.52}$$

and

$$x - \bar{x} = r \frac{s_x}{s_y} (y - \bar{y}) \tag{3.53}$$

The correlation coefficient can be written in the form

$$r = \sqrt{bb'} \tag{3.54}$$

The graphic representation of pairs of regression straight lines showing high and slight degrees of dependence and independence, respectively, is shown in Fig. 3.4.

The value of the correlation coefficient between two variables is influenced by, among other things, autocorrelation in the samples of the random variables, a frequent occurrence in WRS contexts. Therefore, for individual cases under investigation, it is advisable to check the significance of autocorrelation in series of both random variables X and Y in relation to cross-correlation by means of the criterion

$$\alpha = \frac{1}{1 + 2r_1 r'_1 + 2r_2 r'_2 + \ldots} \tag{3.55}$$

where r_1 is the correlation coefficient between $x(t)$ and $x(t+1)$,
r_2 – the correlation coefficient between $x(t)$ and $x(t+2)$,
r'_1 – the correlation coefficient between $y(t)$ and $y(t+1)$,
r'_2 – the correlation coefficient between $y(t)$ and $y(t+2)$, etc.

If α converges to one the cross-correlation coefficient can be calculated as if X and Y were two samples without any autocorrelation (Solomon, 1970). Conversely, the autocorrelation may be analysed in relation to cross-correlation (Matalas, 1970; Cavadias, 1970).

3.1.3.4 Methods of Determination of Complex Statistical Dependence Between Random Variables

Non-linear and multivariate correlation and regression are the main examples of complex statistical dependence, which is important for WRS problems.

Non-linear regression is applied in cases where no other transformation to linear regression is possible. The plotted points are concentrated more or less closely along some curve. There are no simple and exact methods for the determination of this curve, and therefore the form of the regression curve must be estimated. An analogy of the coefficient of linear regression is the *correlation ratio* η and an analogy of the regression straight lines are the regression curves with the *correlation indices* I used as their parameters. They are determined by the equations

$$I_x = \sqrt{1 - \frac{s_{yx}^2}{s_x^2}} \tag{3.56}$$

$$I_y = \sqrt{1 - \frac{s_{xy}^2}{s_y^2}} \tag{3.57}$$

$$\eta = \sqrt{1 - \frac{\bar{s}_y^2}{s_y^2}} \tag{3.58}$$

where s_{xy}^2 and s_{yx}^2 are the residual variances of empirical points with respect to the estimated regression curve and

$$\bar{s}_y^2 = \frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{k} (y_i - \bar{y}_j)^2 \right] \tag{3.59}$$

where k is the number of classes in which the points have been grouped. If $\eta^2 - r^2 \to 0$ the correlation converges on the linear, if $I_x^2 - \eta^2 \to 0$ and also $I_y^2 - \eta^2 \to 0$ the regression function was well estimated. The range of the correlation ratio is the same as the range of the coefficient of linear correlation, i.e. $|\eta| \leqq 1$. In general $|\eta| > |r|$. The difference $\varkappa = \eta^2 - r^2$ is called the *criterion of linearity*.

The transformation of random variables from a non-linear to a linear relationship can have a logarithmic form. This *logarithmic anamorphosis* is very useful in many cases.

In hydrological and other WRS issues multivariate random variables are often investigated (e.g. multivariate regression models). The relationship of the random variables is described by equations of multivariate (in the following example by threevariate) correlation and regression. The regression function in the case of two dependent variables x_1 and x_2 (regression plane) has the following form

$$y = a + b_1 x_1 + b_2 x_2 \tag{3.60}$$

and the expanded form

$$y - \bar{y} = \frac{r_{x_1 y} - r_{x_1 x_2} r_{x_2 y}}{1 - r^2_{x_1 x_2}} \frac{\sigma_y}{\sigma_{x_1}} (x_1 - \bar{x}_1) + \\ + \frac{r_{x_2 y} - r_{x_1 x_2} r_{x_1 y}}{1 - r^2_{x_1 x_2}} \frac{\sigma_y}{\sigma_x} (x_2 - \bar{x}_2) \tag{3.61}$$

with the usual meaning of the symbols used, and the multiple *correlation coefficients* is

$$r_{y.x_1 x_2} = \sqrt{\frac{r^2_{x_1 y} + r^2_{x_2 y} - 2 r_{x_1 x_2} r_{x_1 y} r_{x_2 y}}{1 - r^2_{x_1 x_2}}} \tag{3.62}$$

In this form the relationship between the independent variable Y and two dependent random variables X_1 and X_2 is tested. In hydrological applications the graphic coaxial correlation method is used for the multivariate correlation and regression, e.g. for the rainfall – runoff relationship (Dub and Němec, 1969).

3.1.4 Statistical Estimation of Probability Distribution Parameters

The purpose of the statistical estimation of probability distribution parameters is to provide information for scientific conclusions concerning mass reproducible phenomena and processes, based on samples of observed and measured data. The relationship between the population parameters and the sample estimator is investigated.

3.1.4.1 Estimators of Samples from a Population with Normal Distribution

The population is assumed to be infinite. Random or controlled samples from this population have a finite number of elements (size). The parameters of the population are related to the estimators derived from the sample. These estimators are known functions of the sample elements, which are used in place of the unknown true values

of the estimated parameters. Therefore they vary to some extent from sample to sample. The main parameters and estimators are provided in Table 3.12.

Table 3.12 Main population parameters and sample estimators and their notation

Quantity	Population parameters	Sample estimators
size number of elements	$N\ (N \to \infty)$	n
mean	$\mu(x)$	$\bar{x}$
variance	$\sigma^2(x)$	s_x^2
standard deviation	$\sigma(x)$	s_x
coefficient of variation	$C_v(x)$	C_{vx}
coefficient of skewness	$C_s(x)$	C_{sx}
coefficient of kurtosis	$E(x)$	E_x
k-th moment	$m_k(x)$	$m_{x,k}$
k-th central moment	$M_k(x)$	$M_{x,k}$
mean of sample means	$\mu(\bar{x})$	
variance of sample means	$\sigma^2(\bar{x})$	
mean of sample variances	$\mu(s_x^2)$	
variance of sample variances	$\sigma^2(s_x^2)$	

Assume a population with normal probability distribution. The point and interval estimates of the mean, variance and standard deviation of sample mean and variances are given in Table 3.13.

A limited number of correlation and regression relationships for the other probability distributions of the population are cited in references (Šor, 1965). Their analytical expression is often complicated and it is often approximated by transformations to the normal probability distribution. Methods using the simulation technique are promising for the derivation of estimates from the random series generated.

3.1.4.2 Distribution Involved in Analysis of Variance

The distributions involved in analysis of variance are used in the estimation of population parameters based on random samples from this population.

They include:

– Probability distribution chi-square; this distribution is used, for example, for

Table 3.13 Estimators of a sample from a population with normal probability distribution

Population parameter – quantity	Sample estimator
mean of sample means	$\mu(\bar{x}) = \mu(x)$
variance of sample means	$\sigma^2(\bar{x}) = \dfrac{\sigma^2(x)}{n}$
standard deviation of sample means	$\sigma(\bar{x}) = \dfrac{\sigma(x)}{\sqrt{n}}$
mean of sample variances	$\mu(s_x^2) = \dfrac{n-1}{n}\,\sigma^2(x)$
variances of sample variances	$\sigma^2(s_x^2) = \dfrac{2(n-1)}{n}\,\sigma^4(x)$
interval estimator for sample mean	$\mu(x) - t_p \dfrac{\sigma(x)}{\sqrt{n}} < \bar{x} < \mu(x) + t_p \dfrac{\sigma(x)}{\sqrt{n}}$
interval estimator for sample variance	$\dfrac{\chi^2_{1-p}}{n}\,\sigma^2(x) < s_x^2 < \dfrac{\chi^2_p}{n}\,\sigma^2(x)$
interval estimator for sample standard deviation	$\sqrt{\dfrac{\chi^2_{1-p}}{n}}\,\sigma(x) < s_x < \sqrt{\dfrac{\chi^2_p}{n}}\,\sigma(x)$
interval estimator for sample variance of a large sample	$\sigma^2(x)\left(1 - t_p\sqrt{\dfrac{2}{n}}\right) < s_x^2 < \sigma^2(x)\left(1 + t_p\sqrt{\dfrac{2}{n}}\right)$

χ^2_p, χ^2_{1-p} – tabulated values of a standardized variable having the chi-square probability distribution for probability $p = 1 - \alpha/2$

t_p – tabulated values of the standard variable having the normal distribution for probability $p = 1 - \alpha/2$, where α = level of significance

the sample standard deviation of a random sample from a normally distributed population;

– Student probability distribution; the standardized value of the sample mean if the sample standard deviation is used for this standardization, and also the standardized difference between two different samples from the same population have Student distribution;

– Distribution F (Snedecor); this distribution is the ratio of two different variances of samples from the same population;

– The Fisher distribution is a logarithmic transformation of the Snedecor distribution.

Table 3.14 Distributions involved in analysis of variance

Quantity	Student probability distribution	Chi-square probability distribution	Snedecor probability distribution F
p.d.f.	$f_m(t) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\,\Gamma\left(\frac{m}{2}\right)} \left(1 + \frac{t^2}{m}\right)^{-(m+1)/2}$	$f_m(\chi^2) = \frac{1}{2^{m/2}\,\Gamma\left(\frac{m}{2}\right)} (\chi^2)^{m/2-1}\, e^{-\chi^2/2}$	$f_{m_1 m_2}(F) = \frac{m_1}{m_2} \frac{\Gamma\left(\frac{m_1+m_2}{2}\right)}{\Gamma\left(\frac{m_1}{2}\right)\Gamma\left(\frac{m_2}{2}\right)} \cdot \left(\frac{m_1}{m_2} F\right)^{(m_1/2)-1} \left(1 + \frac{m_1}{m_2} F\right)^{-(m_1+m_2)/2}$
mean	$\mu(t) = 0$	$\mu(\chi^2) = m$	$\mu(F) = \frac{m_2}{m_2 - 2}$
variance	$\sigma^2(t) = \frac{m}{m-2}$	$\sigma^2(\chi^2) = \sqrt{2m}$	$\sigma^2(F) = \frac{m^2}{(m_2-2)^2} \frac{2(m_1+m_2-2)}{m_1(m_2-4)}$
coefficient of skewness	$C_s(t) = 0$	$C_s(\chi^2) = 2\sqrt{\frac{2}{m}}$	
coefficient of kurtosis	$E(t) = \frac{6}{m-4}$		
remarks	$t = \frac{z\sqrt{m}}{\sqrt{v}}$ m – degrees of freedom z – independent random variable with normal probability distribution v – independent random variable with chi-square probability distribution	m – degrees of freedom	$F = \frac{u/m_1}{v/m_2}$ u, v – independent random variables with chi-square probability distribution For $z = \frac{1}{2}\lg F$ Fisher probability distribution m_1, m_2 – degree of freedom for u and v resp.

The p.d.f. and the most important statistical moments of these distributions are given in Table 3.14.

The application of these distributions to point and interval estimation of population parameters and to testing statistical hypotheses is demonstrated in sections 3.1.4.3 and 3.1.5, respectively.

3.1.4.3. Estimators of Population Parameters with Normal Probability Distribution

The estimators are classified as point and interval estimators. In point estimation, the population parameter is estimated by a single value. In interval estimation, two statistics are determined in such a way that the estimated parameter will be inside the interval bounded by them with a given large probability. Interval estimation is, in general, more useful; the range of the interval is related to the accuracy of the estimation.

a) Point estimation

The point estimator of the unknown population mean $\mu(x)$ is the sample mean $\bar{x}$, i.e. $\hat{\mu}(x) = \bar{x}$. The point estimator of the population variance $\sigma^2(x)$ is given by

$$\hat{\sigma}^2(x) = \frac{n}{n-1} s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_1 - \bar{x})^2 \tag{3.63}$$

Since the mean of the sample variances $\mu(s_x^2)$ is equal to

$$\mu(s_x^2) = \frac{n-1}{n} \sigma^2(x) \tag{3.64}$$

the point estimator of the population standard deviation is

$$\hat{\sigma}(x) = \sqrt{\frac{n}{n-1}} s_x^2 = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_1 - \bar{x})^2} \tag{3.65}$$

b) Interval estimation

With probability $p = 1 - \alpha/2$ (called the confidence coefficient; α will be defined in section 3.1.5 as the level of significance), the *population mean* $\mu(x)$ is inside the interval

$$\bar{x} - t_p \frac{\sigma(x)}{\sqrt{n}} < \mu(x) < \bar{x} + t_p \frac{\sigma(x)}{\sqrt{n}} \tag{3.66}$$

It is a two-sided confidence interval. The one-sided confidence interval with probability $p = 1 - \alpha$ is

$$\mu(x) < \bar{x} + t_p \frac{\sigma(x)}{\sqrt{n}} \text{ with lower boundary } -\infty \tag{3.67}$$

or

$$\mu(x) > \bar{x} - t_p \frac{\sigma(x)}{\sqrt{n}} \text{ with upper boundary } +\infty \tag{3.67'}$$

The values t_p are determined from the tables of normal distribution. If the population standard deviation $\sigma(x)$ is not known, it is replaced by its point estimator s_x, and t_p is determined from the tables of the Student distribution for $m = n - 1$ degrees of freedom, where n is the size of the sample from the population. The Student distribution converges asymptotically to the normal distribution for $n \to \infty$.

With probability $p = 1 - \alpha/2$, the population variance $\sigma^2(x)$ is inside the interval

$$\frac{n}{\chi_p^2} s_x^2 < \sigma^2(x) < \frac{n}{\chi_{1-p}^2} s_x^2 \tag{3.68}$$

where χ_p^2 is determined from the tables of the chi-square c.d.f. If the point estimator of variance is known, the following inequalities can be written

$$\frac{n-1}{\chi_p^2} \dot{\sigma}^2(x) < \sigma^2(x) < \frac{n-1}{\chi_{1-p}^2} \dot{\sigma}^2(x) \tag{3.69}$$

Often the one-sided confidence interval is used for the determination of the upper confidence limit of the unknown population variance

$$\sigma^2(x) < \frac{n}{\chi_{1-p}^2} s_x^2 \tag{3.70}$$

or

$$\sigma^2(x) < \frac{n-1}{\chi_{1-p}^2} \dot{\sigma}^2(x) \tag{3.70'}$$

For large samples the confidence interval becomes

$$\frac{s_x^2}{1 + t_p \sqrt{\frac{2}{n}}} < \sigma^2(x) < \frac{s_x^2}{1 - t_p \sqrt{\frac{2}{n}}} \tag{3.71}$$

where t_p is determined from the tables of normal distribution (standardized variable t for probability $p = 1 - \alpha/2$).

The confidence coefficient p used for the estimation of population parameters is often chosen in the interval 95 to 99 % (in exceptional cases 90 %). On the other hand the necessary sample size can be calculated for a given confidence coefficient.

If the acceptable relative error of the mean is given by

$$\vartheta = \frac{t_p \frac{\sigma(x)}{\sqrt{n}}}{\mu(x)} \tag{3.72}$$

then

$$n = \left(\frac{t_p\, C_v(x)}{\vartheta}\right)^2 \tag{3.73}$$

where

$$C_v(x) = \frac{\sigma(x)}{\mu(x)} \tag{3.74}$$

t_p is to be found in tables of normal c.d.f. for probability $p = 1 - \alpha/2$

Example IV

Calculate the interval estimators of the population mean and standard deviation based on the sample of size $n = 280$ with normal probability distribution, sample mean $\bar{x} = 10.37$ and variance $s_x^2 = 0.225$ from example I. The confidence coefficient is 99.5%, the level of significance is 1%.

Solution

1. The confidence interval for $\mu(x)$ is

$$\bar{x} - t_p \frac{\dot{\sigma}(x)}{\sqrt{n}} < \mu(x) < \bar{x} + t_p \frac{\dot{\sigma}(x)}{\sqrt{n}} \tag{IV.1.}$$

where t_p can be found from tables of the Student probability distribution for $m = n - 1$ degrees of freedom. In our case $\alpha = 0.01$; $m = 280 - 1 = 279$; $t_p = 2.594$ for $p = 1 - 0.01/2 = 0.995$.

$\dot{\sigma}(x)$ is the point estimator of the population standard deviation that is calculated as the sample standard deviation with denominator $(n - 1)$ instead of n, i.e.

$$\dot{\sigma}(x) = s_x = 0.475$$

Substituting in (IV.1) we get

$$10.37 - 2.594 \frac{0.475}{\sqrt{280}} < \mu(x) < 10.37 + 2.594 \frac{0.475}{\sqrt{280}}$$

$$10.297 < \mu(x) < 10.444$$

2. The confidence interval of the population standard deviation $\sigma(x)$ is given by inequalities

$$\dot{\sigma}(x) \sqrt{\frac{n-1}{\chi_p^2}} < \sigma(x) < \dot{\sigma}(x) \sqrt{\frac{n-1}{\chi_{1-p}^2}} \tag{IV.2.}$$

For the confidence coefficient $p = 1 - \alpha/2 = 0.995$,

$$0.475 \sqrt{\frac{279}{\chi_{0.995}^2}} < \sigma(x) < 0.475 \sqrt{\frac{279}{\chi_{0.005}^2}}$$

In tables of the chi-square distribution with $m = 279$ degrees of freedom it was found that

$$\chi^2_{0.995} = 343.6 \quad \text{and} \quad \chi^2_{0.005} = 221.9$$

giving

$$0.428 < \sigma(x) < 0.533$$

3.1.5 Tests of Significance and Tests of Hypotheses

The testing of statistical hypotheses involves:

(1) Testing the values of parameters of the probability distribution that is assumed to be known, and

(2) testing the hypothesis of the possibility to fit some probability distribution. A null hypothesis is formulated and its validity is tested with a certain high statistical confidence; i.e. at a given level of significance α, this is the probability (small) of rejecting the null hypothesis when it is true. Its value is selected arbitrarily, e.g. in hydrology and WRS problems values of 1 % and 5 % are mainly used.

Another, more general, possible classification distinguishes between the testing of parametric statistical hypotheses (investigating parameters of the c.d.f. and their sample estimators) and non-parametric ones (investigating other phenomena). The methods of parametric and non-parametric testing may, but in some cases need not, depend on the type of probability distribution.

3.1.5.1 Testing Parametric Hypotheses

Testing parametric hypotheses deals mainly with the problem of decision if a given sample is a random sample from a population whose c.d.f. depends on a parameter; this testing method has been thoroughly investigated for the normal probability distribution:

a) To testing the relationship of a population with mean μ and the random sample with mean $\bar{x}$.

The null hypothesis that the random sample with mean $\bar{x}$ was derived from a normally distributed population with parameters $\mu(x)$ and $\sigma(x)$ is tested. The criterion has the following form

$$t_0 = \frac{|\bar{x} - \mu(x)|}{\sigma(x)} \sqrt{n} \tag{3.75}$$

or

$$t_0 = \frac{|\bar{x} - \mu(x)|}{s_x} \sqrt{n - 1} \tag{3.75'}$$

The null hypothesis is accepted of $t_0 < t_p$; t_p (for probability p) is found in the tables of normal distribution, using the test criterion (3.75) or the criterion (3.75') for $p = 1 - \alpha/2$, i.e. the level of significance is $\alpha = 1 - 2p$.

b) To testing the relationship of the population with variance $\sigma^2(x)$ and the random sample with variance s_x^2.

The test criterion with the probability distribution chi-square has the following form

$$\chi_0^2 = n \frac{s_x^2}{\sigma^2(x)} \tag{3.76}$$

If $\chi_0^2 > \chi_{1-p}^2$ (for $n - 1$ degrees of freedom) with $s_x^2 < \sigma^2(x)$ and $\chi_0^2 < \chi_p^2$ (for $n - 1$ degrees of freedom) with $s_x^2 > \sigma^2(x)$, the null hypothesis is accepted (i.e. the random sample with variance s_x^2 was derived from the normally distributed population with variance $\sigma^2(x)$).

c) Statistical significance of the difference between two sample means.

Two independent random samples of size n_1 and n_2 and with means $\bar{x}_1$ and $\bar{x}_2$, respectively, are from the same population if

$$t_0 = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma(x)} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} < t_p \tag{3.77}$$

or

$$t_0 = \frac{|\bar{x}_1 - \bar{x}_2| \sqrt{n_1 + n_2 - 2}}{\sqrt{n_1 s_{x_1}^2 + n_2 s_{x_2}^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} < t_p \tag{3.78}$$

The value t_p in (3.77) is the 100 p% point of the normal distribution and in (3.78) it is the 100 p% point of the Student distribution at the level of significance $\alpha = 1 - 2p$. If the variances $s_{x_1}^2$ and $s_{x_2}^2$ are clearly different an alternative method of testing is necessary (Votruba and Nacházel, 1969).

d) Statistical significance of the ratio of two sample variances.

Two independent random samples of size n_1 and n_2 with variances $s_{x_1}^2$ and $s_{x_2}^2$ respectively, are from the same population if

$$F_0 = \frac{\dot{\sigma}^2(x_1)}{\dot{\sigma}^2(x_2)} = \frac{s_{x_1}^2}{s_{x_2}^2} \text{ (for } s_{x_1}^2 > s_{x_2}^2) < F_p \tag{3.79}$$

F_p is the 100 . p% point of the Snedecor distribution F and the level of significance is $\alpha = 1 - p$ with $(n_1 - 1)$ and $(n_2 - 1)$ degrees of freedom. A few tests for a population with other than normal probability distribution have been developed.

Example V

Let the hydrological time series of annual flows Q_r at point X for $N = 100$ years be the population with mean $\mu(x) = 200.4$ m^3 s^{-1}. Decide if the deviation of this value from the sample mean $\bar{x}_1 = 196$ m^3 s^{-1} of a 30-year sample from this population is statistically significant. The sample standard deviation is $s_{x_1}^2 = 65$ m^3 s^{-1}.

Solution

Test the null hypothesis that the random sample with mean $\bar{x}_1$ is from the population with mean $\mu(x)$ and standard deviation $\sigma(x)$, which is not known.

The test criterion has the following form

$$t_0 = \frac{|\bar{x}_1 - \mu(x)|}{s_{x_1}} \sqrt{n-1} = \frac{|196 - 200.4|}{65} \sqrt{30-1} = 0.365 \qquad \text{(V.1.)}$$

For different samples with means $\bar{x}_i$ the values t_0 have the Student probability distribution. The level of significance is $\alpha = 1 - 2p = 0.05$, therefore $p = 0.975$. For this p and for $m = n - 1 = 29$ degrees of freedom of the Student probability distribution, the corresponding t_p is

$$t_p = t_{0.975} = 2.045$$

Comparing t_0 with t_p

$$t_0 = 0.365 < t_p = 2.045$$

From this comparison the following conclusion can be drawn. At the selected level of significance $\alpha = 0.05$ the null hypothesis that this sample is from the defined population with normal probability distribution can be accepted. The difference between the sample and population means is not statistically significant.

3.1.5.2 Testing Non-Parametric Hypotheses

The simple estimation of the theoretical probability distribution of a random variable X can be uncertain. The testing of the qualitative and quantitative fit of the theoretical and empirical c.d.f.s without testing their parameters can be performed by the goodness-of-fit criteria included in non-parametric statistical testing methods.

a) Chi-square (Pearson's) criterion

Let us test if the random variable X has the probability distribution of a certain type given by the sample class frequencies. The test criterion is

$$\chi_0^2 = \sum_{i=1}^{l} \frac{(n_i - nf_i)^2}{nf_i} \qquad (3.80)$$

where l is the number of classes,

n – sample size,

n_i – class frequency of i-th class,

f_i – theoretical probability in the case of an accepted null hypothesis (accepted type of probability distribution). For $m = l - c - 1$ (c = number of parameters of the c.d.f. used in the null hypothesis) and level of significance $\alpha = 1 - p$, the value χ_p^2 is found from tables. If $\chi_0^2 < \chi_p^2$ the null hypothesis concerning the goodness-of-fit of the empirical and theoretical probability distributions is accepted.

b) Bernshteyn's criterion

Bernshteyn's criterion is constructed in such a way that the goodness-of-fit is accepted if

$$\frac{1}{l-1} \sum_{i=1}^{l} \frac{(n_i - nf_i)^2}{nf_i} < 1 \tag{3.81}$$

Table VI. 1 Computation of the value of the test criterion

i	Value of class mark x_i	Sample frequencies n_i	Theoretical probabilities of occurrence *) nf_i	Adjustment of the marginal intervals **) nf_i	$n_i - nf_i$	$(n_i - nf_i)^2$	$\frac{(n_i - nf_i)^2}{nf_i}$
1	9.00	1	1.19				
2	9.25	3	4.36	5.55	−1.55	2.40	0.432
3	9.50	11	12.14	12.14	−1.14	1.30	0.107
4	9.75	24	26.00	26.00	−2.00	4.00	0.154
5	10.00	42	43.00	43.00	−1.00	1.00	0.023
6	10.25	62	54.50	54.50	+7.50	56.25	1.032
7	10.50	58	54.30	54.30	+3.70	13.69	0.252
8	10.75	44	42.05	42.05	+1.95	3.80	0.090
9	11.00	19	25.85	25.85	−6.85	47.00	1.820
10	11.25	9	12.15	12.15	−3.15	9.92	0.816
11	11.50	4	3.95				
12	11.75	2	1.10	5.45	−1.55	2.40	0.440
13	12.00	1	0.40				
Σ							5.166

*) The values were transformed by linear interpolation from the values for the lower bound of the values for the lower bound of the class interval theoretical values in Table IV.2. $(nh/s)\, f(t) = nf_i$ to the class mark. As the class length is small, this interpolation will not induce a significant error

**) The marginal intervals were joined in order to reach $nf_i \geqq 5$

c) Kolmogorov–Smirnov test

This is used for testing of the goodness-of-fit of the empirical and theoretical c.d.f.s (or exceedance probability curves). The maximum difference between their values (i.e. between the c.d.f. $F(x_i)$ and the cumulative frequency $F_n(x_i)$) determines the probability

$$P\{\max |F_n(x_i) - F(x_i)| > D_\alpha(n)\} = \alpha \tag{3.82}$$

The values of the test criterion $D_\alpha(n)$ for the sample size n and level of significance α have been tabulated. The null hypothesis for goodness-of-fit of the two functions is accepted if

$$D = \max |F_n(x_1) - F(x_i)| < D_\alpha(n) \tag{3.83}$$

Example VI

The goodness-of-fit between the empirical frequencies in example I that seem to have normal distribution and the theoretical normal probability distribution is to be tested.

Solution

a) Application of the chi-square criterion

Table VI.I. contains the observed frequencies and the corresponding theoretical probabilities of occurrence in the class intervals. The input values in Table VI.I. were taken from Tables I.1 and I.2. The value of the Pearson (chi-square) criterion is

$$\chi_0^2 = \sum_{i=1}^{l} \frac{(n_i - nf_i)^2}{nf_i} = 5.166 \tag{VI.1}$$

The degree of freedom for $c = 3$ are

$$m = l - c - 1 = 10 - 3 - 1 = 6$$

For selected level of significance $\alpha = 0.05$ and one-sided confidence interval with $p = 1 - \alpha$, the following values were found from tables:

$$\chi_{0.95}^2 = 12.6$$

and comparing

$$\chi_0^2 = 5.166 < \chi_{0.95}^2 = 12.6$$

Comparison of these values produces the following conclusion: the null hypothesis concerning the goodness-of-fit of the empirical and theoretical probability distributions at the chosen level of significance can be accepted.

b) Use of Bernshteyn's criterion

Bernshteyn's criterion has the following general form

$$\frac{1}{l-1}\sum_{i=1}^{l}\frac{(n_i - nf_i)^2}{nf_i} < 1 \tag{VI.2}$$

The value of the criterion is

$$\sum_{i=1}^{l}\frac{(n_i - nf_i)^2}{nf_i} = 5.166$$

If the number of the adjusted class interval is $l = 10$, then

$$\frac{1}{10-1}\,5.116 = 0.574 < 1$$

Using this criterion for comparison the same conclusion can be made, i.e. the null hypothesis concerning the goodness-of-fit between the empirical and theoretical probability distributions at the chosen level of significance can be accepted.

3.1.5.3 Testing Hypotheses for Statistical Dependence of Random Variables

a) An infinite population with normal probability distribution has a linear statistical dependence between random variables characterized by the coefficient of linear correlation ϱ. The random samples from it have correlation coefficients r. The test criterion for checking the deviation of r from ϱ has the following form

$$t_0 = \frac{|z - \mu(z)|}{\sigma(z)} \tag{3.84}$$

where z is a function of r and the transformation has the following form

$$z = \frac{1}{2}\lg\frac{1+r}{1-r} \tag{3.85}$$

with parameters

$$\mu(z) \doteq \frac{1}{2}\lg\frac{1+\varrho}{1-\varrho}; \tag{3.86}$$

$$\sigma^2(z) = \frac{1}{n-3} \tag{3.87}$$

for sample size n.

The value t_0 is compared with the critical value t_p for $p = 1 - \alpha$ found in tables of normal probability distribution. If $t_0 < t_p$ the deviation of r from ϱ is random and and the null hypothesis (that this sample with r is from population with ϱ) is accepted.

b) Significance of the difference between two sample correlation coefficients r_1 and r_2 is tested by the criterion

$$t_0 = |z_1 - z_2| \left(\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3} \right)^{-1/2} \tag{3.88}$$

where z_1 and z_2 are transformed values of r_1 and r_2 respectively, using equation (3.85), and n_1 and n_2 are the sample sizes of the correlated pairs. The value t_0 is compared with the value t_p from normal distribution for $p = 1 - \alpha/2$. If $t_0 < t_p$ the null hypothesis of statistical significance between r_1 and r_2 is rejected. Similarly, the sample regression coefficients b_1 and b_2 can be tested by a test criterion with Student probability distribution.

3.2 THEORY OF RANDOM PROCESSES

3.2.1 Basic Notions of the Random Processes Theory

The theory of random (stochastic) processes, as a discipline of probability theory, deals with the investigation of random variables as function of variable parameters, mainly time. Probability theory deals with the probability distribution of one random variable or a finite number of them. On the other hand, the theory of random processes deals with the statistical properties of different (often time related) random functions describing the same physical phenomenon.

Typical examples of random processes in WRS are the time series of river flows, precipitation, temperature, air pressure, and also the states of reservoir storage volumes, etc. Two areas of water management were particularly suitable for the application of the theory of stochastic processes, viz.:

(1) Processing of hydrological data and

(2) probability methods of reservoir design.

The theory of random processes, however, is of wider scope and application. Its methods and procedures (often referred to as stochastic methods and procedures) are used in the investigation of the properties of various random variables in many technical sectors. They are very important in the treatment of WRS problems where together with systems methods, they form the main mathematical research tool[1]).

Quantitatively, the stochastic process is described by a random function of time $X(t)$, the values of which at any time are random variables. The random functions

[1]) Based on the theory of random processes, stochastic hydrology was developed as a subject that investigates the stochastic properties of hydrological processes and constructs mathematical models of them. There is an extensive literature on stochastic hydrology (e.g. included in Loucks *et al.*, 1981; Yevjevich, 1976; Kos, 1982). The development of stochastic hydrology has been promoted at many symposiums and conferences (e.g. Warsaw, 1971, Bratislava, 1975).

are classified with respect to the continuity or discontinuity of the parameter on which they depend: if this parameter is continuous, then the function is called a random (stochastic) process; if it is discrete, then the function is called a random (stochastic) sequence. The methods of their investigation are similar.

There is an analogy between the random variable and the random process: the random variable is described by the set of all its possible values with the probability distribution, the random process $X(t)$ is determined by the set of functions obeying the laws that describe the stochastic properties of this set. The particular functions $x^{(1)}(t)$, $x^{(2)}(t)$ ... are called realization of the stochastic process $X(t)$. In practical applications in WRS the recorded observations in a given time period, i.e. chronological arrangement of random variables, are considered as these realizations.[1])

The probability properties of random processes are investigated, in general, for the whole family of realizations. An idea of such a family is shown in Fig. 3.5 where several possible realizations $x^{(1)}(t)$, $x^{(2)}(t)$, ... are sketched in one graph.

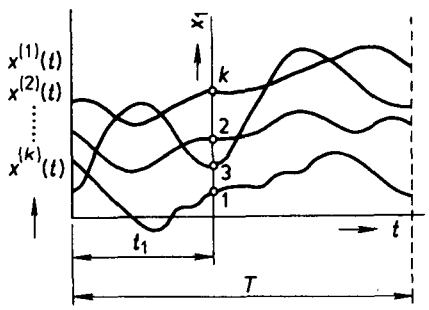


Fig. 3.5 Graph with several realizations of a stochastic process

The stochastic properties of the set (family) of realizations are described by their statistical characteristics (p.d.f., c.d.f. and numerical data, viz. k-th moments and central k-th moments), just as in probability theory these characteristics express the properties of random variables.

The principle difference is the time dependence of statistical characteristics, which is what is used for classification of stochastic processes.

In graphical representation, the straight line $t = t_1$ (perpendicular to the time axis in a fixed point $(t_1, 0)$) crosses the realizations of a set of random variables X_t. If an arbitrary value x_1 is chosen on this line we can investigate the relative frequency $P = m/k$ of occurrence of values $x^{(1)}(t_1), x^{(2)}(t_1), \ldots x^{(m)}(t_1)$ from the total number k of realizations $x^{(k)}(t_1)$ less than x_1. For a large number k of the values of random variable X_1, the relative frequency P converges under presumably "identical" conditions to a constant value for each value x_1 (stochastic stability of frequency P).

[1]) In sampling theory the realization of a random process can be considered as a chronological sample of the population. An arbitrary sample, however, is not a realization.

The c.d.f. of the random variable X_1 is the probability

$$F_1(x_1, t_1) = P(X_1 \leqq x_1) \tag{3.89}$$

which is a function of the selected level x_1 and the time t_1. The subscript of symbol F denotes the first-order (one-dimensional) c.d.f. of the stochastic process which apart from dependence on t_1, does not depend on the further development on the process. An idea of the dynamics of the random process is involved in the investigation of the relative frequency for two times t_1 and t_2 at the chosen levels x_1 and x_2 respectively. In that case, the second-order (two-dimensional) c.d.f. is defined by probability

$$F_2(x_1, t_1; x_2; t_2) = P(X_1 \leqq x_1; X_2 \leqq x_2) \tag{3.90}$$

which is a function of four variables x_1, x_2, t_1, t_2 and gives the probability that for time $t = t_1$ the random variable $X(t)$ is below the level x_1 and also for $t = t_2$ below the level x_2.

In the general case, an n-th order (n-dimensional) c.d.f. is defined as the probability that a random function will be at times $t_1, t_2, \ldots, t_n$ below the values $x_1, x_2, \ldots, x_n$.

The p.d.f.s of the random process are defined as the partial derivatives of the c.d.f. The first-order p.d.f. is defined as the function

$$f_1(x_1, t_1) = \frac{\partial F_1(x_1, t_1)}{\partial x_1} \tag{3.91}$$

and it is called the probability distribution of the random process.

The second-order p.d.f. of the random process is defined as the function

$$f_2(x_1, t_1; x_2, t_2) = \frac{\partial^2 F_2(x_1, t_1; x_2, t_2)}{\partial x_1 \, \partial x_2} \tag{3.92}$$

In general, the n-th order p.d.f. can be defined as the n-th partial derivative of the n-th order c.d.f. with respect to n variables.

The calculation of the c.d.f. and p.d.f. of higher orders is difficult. Consequently, it is often restricted to the calculation of statistical characteristics of the first and second orders. The treatment of many WRS problems often does not require the information provided by the p.d.f. and c.d.f. which would therefore be redundant; the numerical characteristics of the process are frequently sufficient, e.g. the time-related k-th moments of the probability distribution.

The first and second moments are the most important, i.e. the mean value of the random process,

$$m_1\{X(t)\} = \int_{-\infty}^{\infty} x \, f_1(x, t) \, \mathrm{d}x = a(t) \tag{3.93}$$

variance of the random process (second-order central moment)

$$M_2\{X(t)\} = m_2\{[X(t) - a]\}^2 = \int_{-\infty}^{\infty} (x - a)^2 f_1(x, t)\,dx = \sigma^2(t) \tag{3.94}$$

and the autovariance function of the random process (second-order mixed moment)

$$m_1\{X(t_1)\,X(t_2)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2\, f_2(x_1, t_1; x_2, t_2)\,dx_1\,dx_2 = R(t_1, t_2)^1) \tag{3.95}$$

3.2.2 Stationarity and Ergodicity of Random Processes

The stationarity of a random process can be defined by properties of the p.d.f. and c.d.f.

A random process $X(t)$ is stationary if all its p.d.f.s of all orders do not vary with the change of the starting time from t_1 to $t_1 + \tau$. For all values of n and τ,

$$f_n(x_1, t_1; x_2, t_2; \ldots; x_n, t_n) = f_n(x_1, t_1 + \tau; x_2, t_2 + \tau; \ldots; x_n, t_n + \tau) \tag{3.96}$$

Stationarity of a random process can be defined with respect to independence of the c.d.f. of all orders on the time origin. Therefore, if

$$F_n(x_1, t_1; x_2, t_2; \ldots; x_n, t_n) = F_n(x_1, t_1 + \tau; x_2, t_2 + \tau; \ldots; x_n, t_n + \tau) \tag{3.97}$$

for any n and τ the process is stationary.

Proving the stationarity of a random process in WRS contexts using conditions (3.96) and (3.97) involves at intractable calculations. Therefore, the time independence of statistical parameters of the c.d.f. is considered satisfactory and the dependence of the autovariance (autocorrelation) function on the time lag τ only is considered as a single time factor.

Less strict requirements are used in weakly stationary random processes. Unlike strictly stationary random processes which have to satisfy conditions (3.96) or (3.97), a weakly stationary process has to meet the following conditions: (1) the mean value is a constant (time invariant), (2) the autovariance (autocorrelation) function is dependent on the time lag τ only (the probability distributions have the same covariance matrix for all τ). A strictly stationary random process with a finite covariance matrix is also weakly stationary. The converse, however, need not be true.

Random processes that do not satisfy the conditions of stationarity are non-stationary. Fluctuations of statistical characteristics during the development of the stochastic process are typical of them. Their investigation is, therefore, substantially more difficult. The theory of non-stationary processes developed later than the

[1]) Apart from the moments (3.93), (3.94), and (3.95), WRS problems need the coefficients of variance $C_v(t)$ and skewness $C_s(t)$ required by Pearson's probability distribution (type III) of the random process. The meaning and method of calculation of the autocorrelation function is explained in section 3.2.3.

theory of stationary processes. Therefore, an assumption of stationarity in hydrological processes was applied in WRS investigations. However, non-stationary hydrological processes, their probability properties and their effect on WRS are now being investigated (Nacházel *et al.*, 1975, Yevjevich, 1976).

The second important property of random processes that is related to stationarity is ergodicity. A random process is ergodic if each statistical characteristic calculated as the mean from set of possible realizations can be obtained (with probability converging to one) from one realization of the stochastic process using a sufficiently long time interval (i.e. a single realization yields sufficient information about the process).

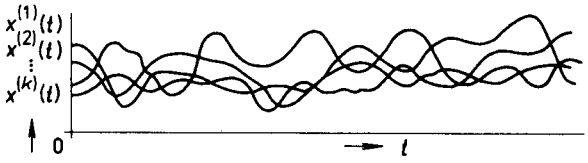


Fig. 3.6 Several realizations of a stationary ergodic random process

Ergodicity and stationarity are two different properties of random processes. Every ergodic process is stationary; however, a stationary process need not necessarily be ergodic. Figure 3.6 contains an example of a stationary ergodic random process with realizations that have a constant mean value in time. Figure 3.7 illustrates an example of a random process with realizations that have different mean values in time. Sufficient for the ergodicity of a stationary process is satisfaction of the requirement of a zero value of the time limit of the autocorrelation function for $\tau \to \infty$ (for definition of the term autocorrelation function, see section 3.2.3).

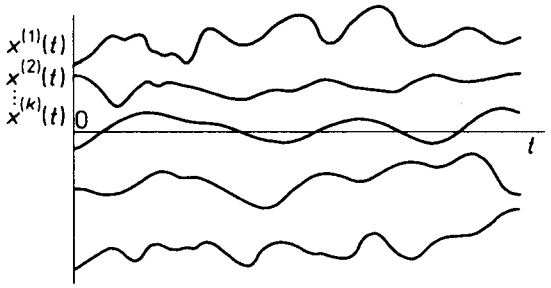


Fig. 3.7 Several realizations of a stationary nonergodic random process

The definition of ergodicity implies that each statistical characteristic of a random process can be obtained as the mean value of a family of realizations or as the mean value of a family of realizations or as the mean value of a single realization in time. The equivalence of these two methods of calculation of basic characteristics is expressed in Table 3.15.

Table 3.15 Statistical characteristics for stationary ergodic random processes

Characteristics	Mean value in time	Mean value of family of realizations
mean	$\bar{x}(t) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} x(t)\,dt$	$a(t) = \int_{-\infty}^{\infty} x\, f_1(x, t)\,dx$
variance	$\overline{x^2(t)} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} x^2(t)\,dt$	$m_2[X^2(t)] = \int_{-\infty}^{\infty} x^2\, f_1(x, t)\,dx$
covariance function	$\overline{x(t)\,x(t+\tau)} = \lim_{T\to\infty} \int_{-T}^{+T} x(t)\,x(t+\tau)\,dt$	$m_1[X(t)\,X(t+\tau)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_2(x_1, t_1; x_2, t_2)\,dx_1\,dx_2$

The usefulness of ergodicity of stationary processes is related to the possibility of investigating a single realization during a long time interval instead of a large set of realizations which often are not available. This property of random processes, like that of stationarity, significantly facilitates the treatment of WRS problems. However, the direct observation of hydrological phenomena is limited in time which justifies a certain lack of confidence in the equality of the mean value determined from a set of realizations and the mean value in time.

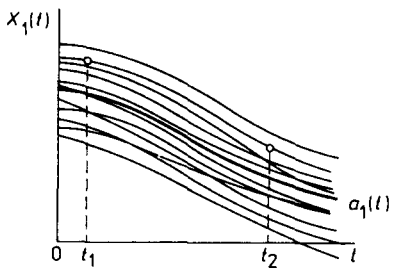


Fig. 3.8 Inner structure of realizations of a random process – regular tendency

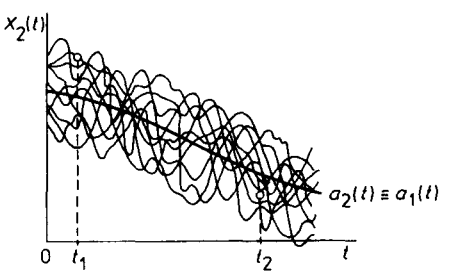


Fig. 3.9 Inner structure of realizations of a random process – irregular tendencies

3.2.3 Correlation Analysis of Random Processes

The full description of a stochastic process includes not only the p.d.f. and c.d.f. and their characteristics but also the results of correlation analysis. Figures 3.8 and 3.9 contain examples of two random processes $X_1(t)$ and $X_2(t)$, each with several realizations having the same means and variances.

The inner structure of random processes is remarkably different. For its identification the autocovariance and autocorrelation functions have been introduced to describe the correlation between the values of the process at any two times t_1 and t_2. The value of the autocorrelation function for the two given times is equal to the correlation coefficient of the corresponding random variables[1]).

The autocovariance function has been defined in eq. (3.95) and Table 3.15 shows the mean values in time and the mean values of the family of realizations. In WRS problems the covariance function of discrete random sequences is frequently calculated from the covariances of random variables at times t_1 and t_2 by the formula

$$R(t_1, t_2) = m_1\{X(t_1)\, X(t_2)\} \tag{3.98}$$

where m_1 is the mean value of the product $X(t_1)\, X(t_2)$.

The transformation from the covariance function (3.98) to the central covariance function (mixed central moments) is given by

$$\mathring{R}(t_1, t_2) = m_1[\mathring{X}(t_1)\, \mathring{X}(t_2)] = m_1\{[X(t_1) - a(t_1)]\, [X(t_2 - a(t_2)]\} \tag{3.99}$$

Dividing the values of the central covariance functions from eq. (3.99) by the product of the standard deviations $\sigma[X(t_1)]\, \sigma[X(t_2)]$ of the corresponding random variables $X(t_1)$ and $X(t_2)$, the normalized covariance function, i.e. correlation function of the random process is

$$r(t_1, t_2) = \frac{\mathring{R}(t_1, t_2)}{\sigma[X(t_1)]\, \sigma[X(t_2)]} = \frac{R(t_1, t_2) - a(t_1)\, a(t_2)}{\sigma[X(t_1)]\, \sigma[X(t_2)]} \tag{3.100}$$

The values of the correlation function are the correlation coefficients between the random variables $X(t_1)$ and $X(t_2)$. It can be shown that for $t_1 = t_2$ the value of the central covariance function is equal to the variance, i.e.

$$\mathring{R}(t_1, t_1) = \sigma^2[X(t_1)] \tag{3.101}$$

and the correlation function is equal to one

$$r(t_1, t_1) = \frac{\sigma^2[X(t_1)]}{\{\sigma[X(t_1)]\}^2} = 1 \tag{3.102}$$

The principal properties of the coefficient of correlation imply that the maximum value of the correlation function is one for $t_1 = t_2$, i.e.

$$|r(t_1, t_2)| \leqq 1 \tag{3.103}$$

The definition of the correlation coefficient implies positive and negative values of the correlation function. If the time lag between the values $X(t_1)$ and $X(t_2)$ is small

[1]) Apart from the autocorrelation function there are other types of correlation function which, however, have other meanings (see Beneš, 1970). When there is no danger of ambiquity, the simpler term correlation function is used.

(i.e. $t_2 - t_1 \to 0$), a close, positive correlation can be expected. For a greater time lag, the correlation function can assume different values according to the degree of dependence or independence. In any case, the values of the correlation function should be statistically tested to distinguish the statistically significant values and random fluctuations.

The correlation function is an important index for stationary and ergodic processes. The covariance and correlation functions of a stationary random process are functions of a single argument, viz. the time lag $\tau = t_2 - t_1$, i.e. for the covariance function

$$R(t_1, t_2) = R(\tau) \tag{3.104}$$

and for the correlation function

$$r(\tau) = \frac{R(\tau) - a^2}{\sigma_x^2} = \frac{\mathring{R}(\tau)}{\sigma_x^2} \tag{3.105}$$

The covariance function is an even function and hence

$$R(\tau) = R(-\tau) \tag{3.106}$$

The same relationship holds for the correlation function[1]).

If the process is stationary and ergodic its correlation function is equal to the correlation function of this realization.

A typical characteristic of a stationary random process is the decrease in interrelatedness between the random variables with increasing time lag τ, until, at the limit $\tau \to \infty$, these variables are independent. Independence can be defined, for example, in this way: for $\tau > \tau_0$ the absolute value of the correlation coefficient is less than some given number, e.g.

$$|r(\tau)| < 0.05 \tag{3.107}$$

where τ_0 is the time of the duration of the correlation.

---

[1]) These properties of covariance and correlation functions are important in practical applications. If the correlation function depends on a single argument τ, then it expresses the relationship of random variables with the time lag τ irrespective of the time origin. As the correlations function is even, the calculation can be performed with either a positive or a negative time lag.

However, the correlation function of some special random processes and functions (e.g. a stationary process with a periodic component, a non-stationary process, or periodic functions of various types) need not necessarily converge. Their correlation analysis is based on the fact that the correlation function of a periodic function is also a periodic function with the same period as the given function and independent of the phase.

In fact, there are various types of correlation functions; they are distinguished by their form, duration, convergence or divergence, periodicity or statistical significance of individual members, all of which describe the nature of the random process and the consequences for further processing.

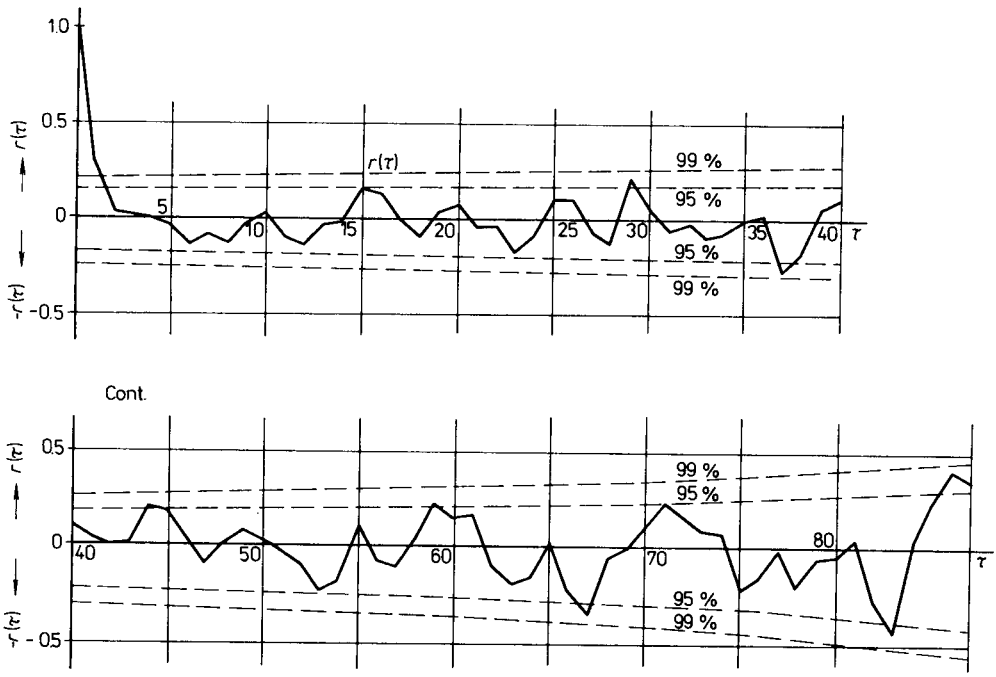


Fig. 3.10 Correlation function of the annual flows of the Labe River in the town of Děčín at time period 1851 – 1957

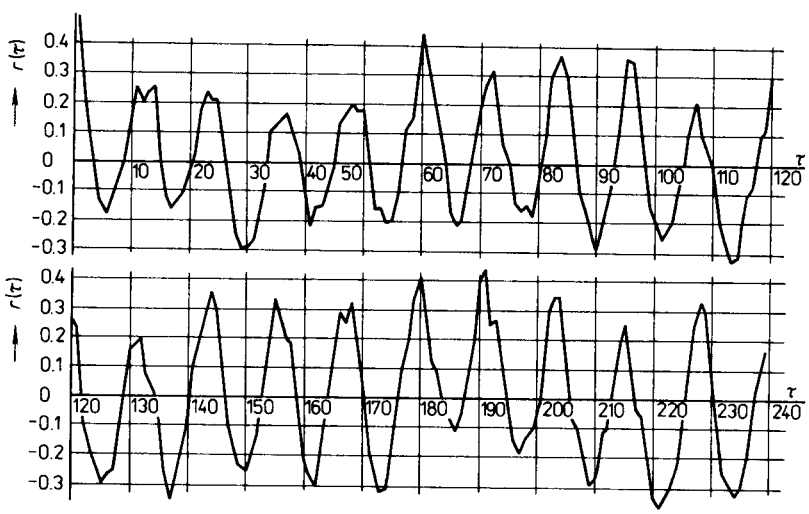


Fig. 3.11 Correlation function of the monthly flows in the Labe River in the town of Děčín at time period 1851 – 1960

If a series of discrete random variables is given in a certain observation period, the correlation function can be calculated in two ways:

(1) The requirement to use a constant number of pairs of random variables results in the calculation of the correlation function from its definition, i.e. the formula

$$r(\tau) = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_i)(x_{i+\tau} - \bar{x}_{i+\tau})}{\sigma_i \sigma_{i+\tau}(n-1)} \tag{3.108}$$

Figure 3.10 contains the correlation function of the annual flows of the River Labe in the town of Děčín for the period 1851 – 1957. It is periodic in form, which is typical of many hydrological patterns. In a long-term analysis, a period of 13 to 15 years can be clearly distinguished and proved by special analysis (see 3.2.4).

Figure 3.11 shows the correlation function of the monthly flows of the River Labe in the town of Děčín for the period 1851 – 1960. The correlation function is again periodic, but it has a pronounced annual cycle of flow variation.

The long-term variations in the activity of the sun are important factors in the analysis of geophysical and hydrological processes and their interrelations. Figure 3.12 contains the correlation function of the mean annual relative Wolf numbers for the period 1831 – 1894 (the calculation was carried out up to time lag $\tau = 35$). The pattern of this line is harmonic and regular, showing the well-known 11-year cycle of the sun's activity. Compared with the correlation functions of the flows or rainfall time series, in this case the correlation relationship is very close, for direct or indirect correlation. The periodicity is obvious and significant.

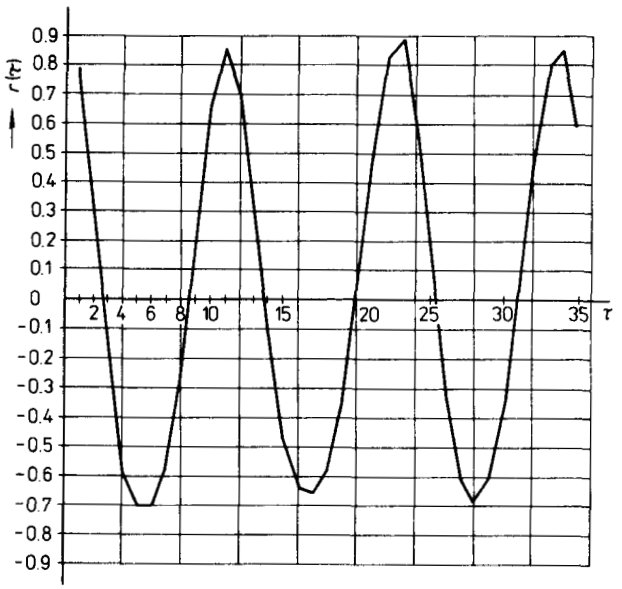


Fig. 3.12 Correlation function of the mean annual relative Wolf's numbers at time period 1831 – 1894

where x_i are random variables with values x_1 to x_n,
$x_{i+\tau}$ – random variables with values $x_{1+\tau}$ to $x_{n+\tau}$,
τ – the time lag argument of the correlation function, for $\tau = 0$ to $\tau = m - n$ where m is the total number of elements in the series (length of observation),
n – the chosen sample section of the series,
$\bar{x}_i, \bar{x}_{i+\tau}$ – the mean values of random variables x_i and $x_{i+\tau}$ respectively, and
$\sigma_i, \sigma_{i+\tau}$ – the standard deviations of these series.

The advantage of calculating the correlation function from formula (3.108) is that all the correlation coefficients have the same weight, which simplifies the test of statistical significance. The confidence limits are determined by the approximative formula (Anderson, 1942)

$$r_p = \frac{-1 \pm t_p \sqrt{n-2}}{n-1}$$

(for another method of calculation see (3.84) to (3.87)) where t_p is the standardized normal variable corresponding to the level of significance α and $p = 1 - \alpha$. For the usual levels of significance in hydrology viz. 1% and 5%, $t_{0.99} = 2.326$ and $t_{0.95} = 1.645$.

(2) A disadvantage attached to formula (3.108) is that the correlation coefficients do not include all the terms of the series and some information is lost. Therefore the correlation function can be calculated by

$$r(\tau) = \frac{\sum_{i=1}^{m-\tau} (x_i - \bar{x}_i)(x_{i+\tau} - \bar{x}_{i+\tau})}{\sigma_i \sigma_{i+t}(m - \tau - 1)} \tag{3.110}$$

with the same meaning of the variables as in (1).

The disadvantage of the second method of calculation is the diminishing number of pairs in correlation with increasing the time lag τ; the confidence of the values of the correlation function decreases too. Anderson's test of statistical significance takes this fact into consideration and the confidence limits are calculated by

$$r_p = \frac{-1 \pm t_p \sqrt{m - \tau - 2}}{m - \tau - 1} \tag{3.111}$$

where m, t_p has the same meaning as in formula (3.109)[1]). The autocovariance and autocorrelation functions are important for the investigation of correlation properties of one random process; the correlation interrelations of several random processes

[1]) The confidence limits for correlation coefficients calculated by (3.109) are constant; if calculated by (3.111) they are a function of time lag τ and their range increases, which expresses a stricter condition for the statistical significance of the correlation function.

are investigated by cross-covariance and cross-correlation functions. The simplest of these is the cross-covariance function of two random processes, which is defined in the following way:

$$R_{xy}(t_1, t_2) = m_1[X(t_1)\, Y(t_2)] \tag{3.112}$$

$$R_{yx}(t_1, t_2) = m_1[Y(t_1)\, X(t_2)] \tag{3.113}$$

It is also

$$R_{xy}(t_1, t_2) = R_{yx}(t_2, t_1) \tag{3.114}$$

The cross-covariances can also be calculated as the central and normalized moments when the cross-correlation function is obtained in the following form

$$r_{xy}(t_1, t_2) = \frac{R_{xy}(t_1, t_2) - m_x(t_1)\, m_y(t_2)}{\sigma[X(t_1)]\, \sigma[Y(t_2)]} \tag{3.115}$$

$$r_{yx}(t_1, t_2) = \frac{R_{yx}(t_1, t_2) - m_y(t_1)\, m_x(t_2)}{\sigma[Y(t_1)]\, \sigma[X(t_2)]} \tag{3.116}$$

The pattern of the cross-correlation function indicates the degree of the correlation relationship (coherency) between the realizations of the random process with some time lag.

There are many types of cross-correlation functions similar to the autocorrelation function, and they can be distinguished by the nature of the individual random processes. Their investigation is the subject of specialist literature (Beneš, 1961; Levin, 1965; Ventcel, 1964).

3.2.4 Spectral and Filter Analysis of Random Processes

The spectral distribution function and the spectral density function are further important statistical characteristics of a random process or its realization, particularly in the investigation of periodic properties. The spectral density function (s.d.f.) of a random process $X(t)$ is defined by the expression (Beneš, 1961)

$$s_{xx}(\omega) = \lim_{T \to \infty} M \left[\frac{1}{2T} |X_T(\mathrm{i}\omega)|^2 \right] \tag{3.117}$$

where $X_T(\mathrm{i}\omega)$ is the Fourier transform of the random process $X(t)$ in a finite interval $\langle -T, T \rangle$ (see section 3.4), and is given by

$$X_T(\mathrm{i}\omega) = \int_{-\infty}^{\infty} X(t)\, \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}t \tag{3.118}$$

where $X(t)$ is defined by the interval $\langle -T, T \rangle$.

The spectral density function (s.d.f.) is, in fact, a limit of expectation in a set of

of periodograms for $T \to \infty$; the periodogram of a realization of a random process is a Fourier transform[1]).

The s.d.f. (spectral density function) of a random process can be calculated from its autocorrelation function. Wiener and Khinchin have shown that between the autocorrelation function of a stationary process and its s.d.f. the following relationships apply:

$$\mathring{R}(\tau) = \int_0^\infty S_x(\omega) \cos \omega\tau \, d\omega \tag{3.119}$$

$$S_x(\omega) = \frac{2}{\pi} \int_0^\infty \mathring{R}(\tau) \cos \omega\tau \, d\tau \tag{3.120}$$

Equations (3.119) and (3.120), called Wiener–Khinchin equations, express fundamental and very important relationships as they facilitate the transformation from one set of statistical characteristics to a second one (the s.d.f. is a Fourier transform of the autocorrelation function and, conversely, the autocorrelation function is an inverse Fourier transform of the s.d.f.).

If, analogously to the normalized autocovariance function, the normalized s.d.f. is introduced

$$s_x(\omega) = \frac{S_x(\omega)}{\sigma_x^2} \tag{3.121}$$

then equations (3.119) and (3.120) can be arranged in the form

$$r(\tau) = \int_0^\infty s_x(\omega) \cos \omega\tau \, d\omega \tag{3.122}$$

$$s_x(\omega) = \frac{2}{\pi} \int_0^\infty r(\tau) \cos \omega\tau \, d\tau \tag{3.123}$$

In some technical applications it is mathematically easier to express the spectral and correlation functions by complex variables. Substituting in equations (3.122) and (3.123)

$$\cos \omega\tau = \frac{e^{i\omega\tau} + e^{-i\omega\tau}}{2}$$

and letting $s_x(\omega) = 2s_x^*(\omega)$ and increasing the range to $(-\infty, \infty)$, we get

$$r(\tau) = \int_{-\infty}^\infty s_x^*(\omega) \, e^{i\omega\tau} \, d\omega \tag{3.124}$$

[1]) The investigation of periodic properties of random processes using their spectra (spectral representation) is similar to harmonic analysis, which is used for the investigation of non-random processes. Spectral analysis is a generalization of this by introducing the time expectation of spectral representations obtained from realizations.

$$s_x^*(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} r(\tau)\, e^{-i\omega\tau}\, d\tau \tag{3.125}$$

The s.d.f. $s_x(\omega)$ differs from $s_x^*(\omega)$ by different scales on ordinates and by $s_x(\omega)$ not being defined for negative frequencies (Fig. 3.13).

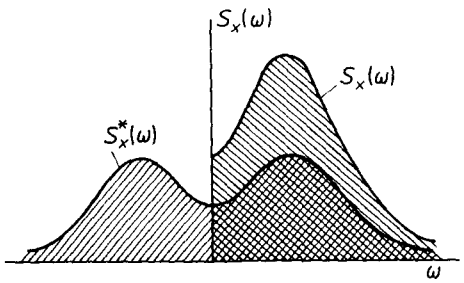


Fig. 3.13 The complex representation of the s.d.f.

Just as the pattern of correlation functions indicates the correlation tendencies in the random process, the pattern of the s.d.f. shows its periodic properties. The limiting cases with characteristic patterns of the s.d.f. are used as the basis for this analysis. For example, a periodic function with the frequency ω_0 has an infinite value of the s.d.f. for the value ω_0 (Fig. 3.14a). This limiting cases, however, is not possible in an actual stochastic process in actual physical conditions. A stochastic process with pronounced periodicity and with a periodic correlation function has the s.d.f. in a narrow frequency band concentrated around the given frequency ω_0 where the maximum occurs (Fig. 3.14b).

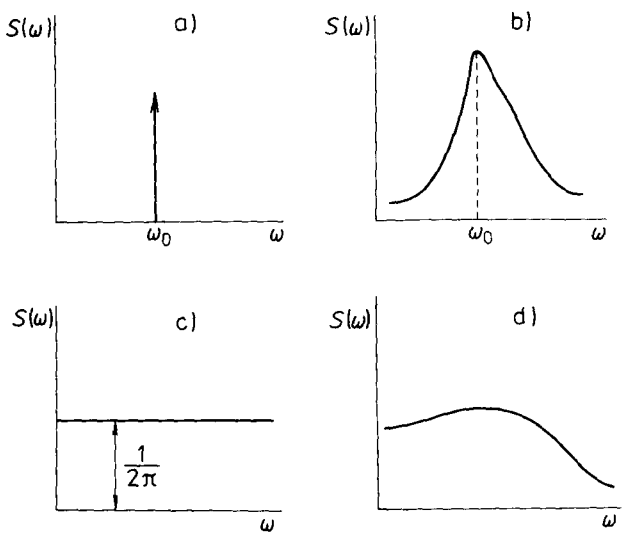


Fig. 3.14 Graphs of s.d.f. *a* – periodic function, *b* – narrow band spectrum process, *c* – white noise, *d* – broad band spectrum process

The opposite case is the s.d.f. of a random process that is constant for all frequencies (Fig. 3.14c). This process is called white noise and its values are uncorrelated for any time lag τ (with the exception of $\tau = 0$). Therefore its correlation function is

$$r(\tau) = \begin{cases} 1 & \text{for } \tau = 0 \\ 0 & \text{for } \tau \neq 0 \end{cases} \tag{3.126}$$

However, white noise is also an ideal process that does not exist in natural conditions since the random functions with a small time lag are in almost all cases dependent. In real processes which converge to white noise a broad-band spectrum occurs (Fig. 3.14d), which can be considered constant in a given interval of frequencies.

White noise has a uniquely defined correlation function and s.d.f. However, they do not uniquely define the probability distribution of the random process. Therefore, white noise is an example of a group of random processes with the same correlation function and s.d.f. but with different probability distributions (Levin, 1965).

Since the correlation function and s.d.f. are two statistic characteristics related by the Fourier transformation, this relationship is often used in hydrology. Figure 3.15

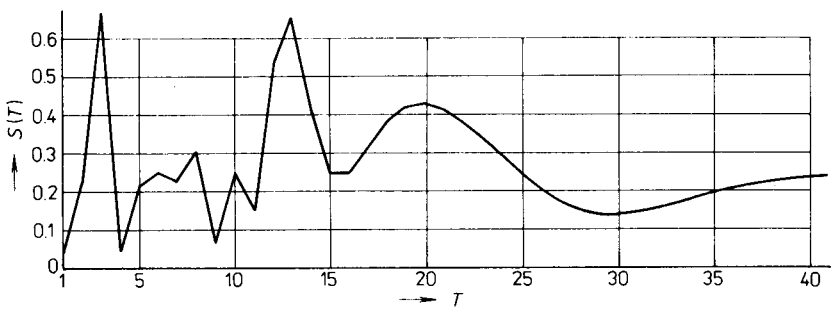


Fig. 3.15 Graph of the s.d.f. of the annual flow time series in the Labe River in the town of Děčín in the period 1851 – 1957

contains an example of the s.d.f. of the annual flows of the River Labe in the town of Děčín. In the graph, the most important extreme values are reached for $T = 3$ years (range of short-term periods) and for $T = 13$ years (range of medium-term periods). The maximum values for $T = 20$ and $T = 42$ years are less pronounced and for $T > 35$ the spectrum has a broad-band character.

If a series of discrete random values is given, its s.d.f. can be calculated by the formula (Reznikovsky, 1969)

$$S = \frac{1}{2\pi}\left[1 + 2\sum_{\tau=1}^{m}\left(1 - \frac{\tau}{m+1}\right) r(\tau)\cos\frac{2\pi\tau}{T}\right] \tag{3.127}$$

where m should be limited by $m = n/2$ (n is the length of the input series). T is the number of years related to the frequency by

$$T = \frac{2\pi}{\omega}$$

In expression (3.127) the various weights of the correlation coefficients of the correlation function are taken into account (3.110).

If the length of the input hydrological time series is limited, the periodic properties determined by spectral analysis need not be statistically significant. The s.d.f. has to be statistically tested; the numerical method of treatment is described in the references (Reznikovsky, 1969; Votruba and Nacházel, 1971).

The theory of the spectral analysis of random processes was developed for stationary stochastic processes. It can be generalized by the Fourier transformation of n-dimensional correlation functions to allow treatment of problems of non-stationary random processes (Batkov, 1959; Lebedyev, 1959; Leonov, 1956; Levin, 1965; Pugachev, 1957).

The method of stochastic filters is a useful complementary method of time series statistical analysis which involves the investigation of their periodic properties. Filters are used for the estimation of random variables, the random components being detected in a given random sequence and then separated. The underlying idea is the concept of the realization of a random sequence as the sum of deterministic variables and random variables in the form of white noise. By special algorithms (filters) the two sequences are separated, which facilitates the probability analysis of the input series.

The method of moving averages (smoothing of the sequence) can also be regarded as a method of filters. The aim is the calculation of a filtered sequence with statistical characteristics (viz. correlation function and s.d.f.) that can be more reliably used for the determination of the nature of the sequence. That is the main practical purpose of the methods of filters (Nacházel and Patera, 1974).[1])

3.2.5 Markov Processes

White noise cannot be applied in all technical problems as the random variables are often stochastically dependent. Hydrological processes are typical correlated random processes that have a great variety of properties. The investigation of random processes with a general type of correlation function is often difficult because the

[1]) In Fig. 3.16 there is an example of the s.d.f. of annual flow time series in the River Labe in the town of Děčín, filtered by binominal filters (Nacházel and Patera, 1974) in three alternatives. Their parameters are interesting in that they indicate the medium-term period T=15 years as the most significant one for all chosen degrees of the filter. This is a good fit with the estimated period of the correlation function of the observed flow series.

development and history of the process (its inner structure) has to be followed in large, theoretically infinite, time intervals. The idea of following the history of a random process in a limited time interval led to the development of random processes of special types, the theory being based on the work of the Russian mathematician A. A. Markov (1856–1922).[1])

If a Markov process is at time t_k in state[2]) $X(t_k)$, then the probability of occurrence of some future state depends only on $X(t_k)$ and not on the preceding history, i.e. not on states $X(t_i)$, $X(t_j)$ where $i < j < k$.

Markov processes can be classified similarly to general random processes using the probability distribution and continuity of the time argument. In WRS issues Markov processes with discrete time (Markov chain) are frequently applied.

Markov chains are described by the transition probabilities, i.e. by conditional probabilities $p_{i,j}$ of moving from state X_i at one time to some state X_j at the following time. If the process can assume a finite number of states $X_1, X_2, \ldots, X_n$, then for each state X_i $(i = 1, 2, \ldots, n)$ at time t_k there are n probabilities P_{ij} $(j = 1, 2, \ldots, n)$ of moving from state X_i to state X_j at time t_{k+1}. The total number of conditional transition probabilities is n^2; they are arranged in a square matrix $(n \cdot n)$, called a transition matrix, which has the following form

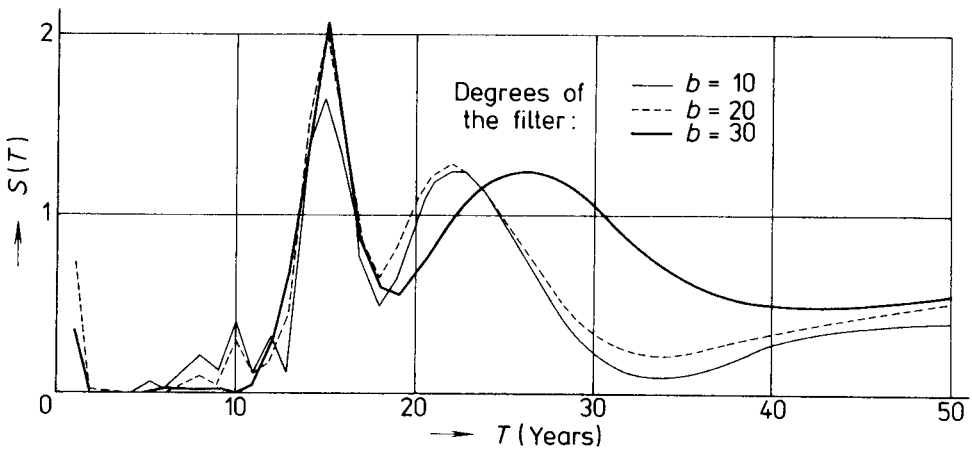


Fig. 3.16 Graph of the s.d.f. of the filtered annual flow time series in the Labe River in the town of Děčín

[1]) The theory of Markov processes is now the principal section of the theory of random processes (Dupač, 1955; Dupač and Dupačová, 1975; Dynkin, 1963). In WRS problems, Markov processes have mainly been used in mathematical models of the hydrological processes and in the probability theory of reservoir release operation. Some WRS optimization methods are likewise based on these principles.

[2]) The state is defined as the numerical value that the random variable $X(t)$ can assume.

$$\mathbf{p}_{ij} = \begin{matrix} p_{11}, & p_{12}, & \ldots, & p_{1n} \\ p_{21}, & p_{22}, & \ldots, & p_{2n} \\ \vdots & & & \\ p_{n1}, & p_{n2}, & \ldots, & p_{nn} \end{matrix} \tag{3.128}$$

Each row vector of this matrix represents the transition probabilities of moving from an initial state to all terminal states and, therefore, the sum of the values of the elements in each row has to be equal to one

$$\left(\sum_{j=1}^{n} p_{ij} = 1 \quad \text{for } i = 1, 2, \ldots, n \right).$$

Each column vector represents the probabilities of moving from different initial states to a given terminal state; the sum of the values of the elements in each column can be different.

The transition probabilities of a Markov chain can be schematically represented (Fig. 3.17) or a digraph can be used (Fig. 3.18) where the nodes of the graph represent

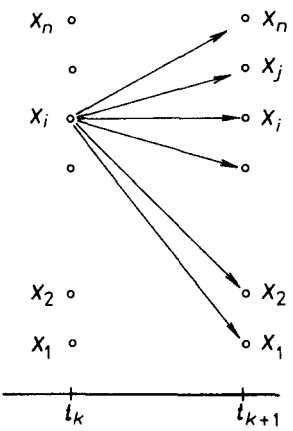


Fig. 3.17 Schematic representation of transition probabilities

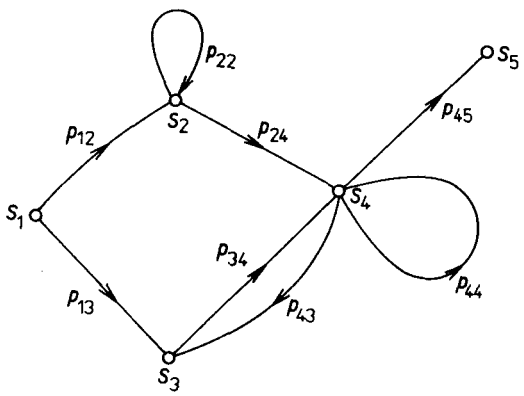


Fig. 3.18 Digraph of transition probabilities

the states of the Markov chain and the arcs of graph the transitions between the states. The missing arcs of the graph of the transition matrix $5 \cdot 5$ indicate the zero value of the corresponding element of the transition matrix (Beneš, 1970).

The Markov process is determined by the transition probabilities from time t_k to times t_{k+1}, t_{k+2}, ... as indicated by its definition. If this relationship is confined to

only two, successive times, the process is defined by a single transition matrix, i.e. it is a first-order (simple) Markov process. If n relationships are taken into consideration, compound Markov processes (or chains) or n-th order Markov processes (chains) are formed.

The form of the Markov chain is often the basis of the mathematical models of hydrological sequences. The hydrological time series is treated as a realization of a compound Markov chain, where some correlation relation occurs not only between consecutive elements of the sequence but also between the elements with time lag 2, 3, ..., τ. The conditional transition probabilities express the probability of occurrence of a random variable Q_i on condition that, at the preceding times, the random variables were equal to

$$Q_{i-1},\ Q_{i-2},\ \ldots,\ Q_{i-\tau}$$

which can be symbolically described as

$$p(Q_i/Q_{i-1},\ Q_{i-2},\ \ldots,\ Q_{i-\tau})$$

The corresponding conditional p.d.f. is

$$f(Q_i/Q_{i-1},\ Q_{i-2},\ \ldots,\ Q_{i-\tau})$$

are the conditional c.d.f. is

$$F(Q_i/Q_{i-1},\ Q_{i-2},\ \ldots,\ Q_{i-\tau})$$

which is used directly for the generation of the random variable Q_i. The conditional c.d.f. is determined by moments in the same way as the unconditional c.d.f.

The construction of mathematical models of random sequences is easier and more accurate with normal probability distribution. In this case, an extensive knowledge of the properties of normal probability distribution related to stochastic processes can be utilized and the assumption of a compound Markov chain can be related to multivariate regression analysis (Hájek and Anděl, 1969).

If the c.d.f.s are skew, some functions are sought for transformation to the normal c.d.f. or some other special methods are used. Experience has shown that these numerical methods often affect the confidence of the mathematical model, and therefore this fact has to be taken into consideration in the evaluation of the deviations of the model output statistical parameters (Nacházel *et al.*, 1975).

3.2.6 Mathematical Models of Stochastic Processes and Their Application in WRS

The importance of modelling of stochastic process in various technical tasks is a consequence of the development of the theory of stochastic processes. Their analysis was considered complete if their characteristics were expressed analytically. The analytical solution was regarded as a further stage of investigation regardless of the mathematical difficulties and degree of tractability. Research development has been influenced principally by the possibility of mathematical modelling of random process realizations of arbitrary length and of using these realizations without solution of the analytical problem.

Modelling of stochastic processes is widely used for WRS problems. The stochastic hydrological time series facilitate the investigation of many aspects of the behaviour of hydrological systems and, in addition, they form a basis for the solution of WRS problems, including hydroelectric power systems, that otherwise could only be partially solved, if at all.

This advantage, which also includes a higher degree of confidence in the calculation of water supply reliability, is a major one and therefore future research into these methods of modelling stochastic hydrological time series will centre on the theory of reservoir storage and the theory of WRS optimization[1]).

In theory, mathematical models of stochastic processes are predominantly based on the assumption of a regression relationship among the variables, on their periodic properties or on a combination of both these principal attributes of stochastic processes. Regression models have now been developed to such a degree that their principles have been successfully applied to monthly hydrological time series and the accompanying systems.

3.2.6.1 Model for Generation of Sequences of Annual Stochastic Hydrological Variables

Applying the results of section 3.2.5, construction of the mathematical model of annual values of random variables (flows, precipitation, etc.) is based on the assumption that there is a linear regression relationship between the value generated at time t and the given values at preceding times (years) $t-1, t-2, \ldots, t-k$. This relationship reflects the probability properties of the sequence observed (re-

[1]) That is why such attention is being paid to the development of stochastic process modelling. In Czechoslovakia, remarkable progress was achieved by, e.g., Votruba and Broža, 1966; Kos, 1969; Votruba and Nacházel, 1971; Nacházel and Bureš, 1973; Nacházel *et al.*, 1975. These references include a description of the development of these methods and a detailed explanation of them.

alization) and the number of preceding times k is determined by the order of the Markov chain. The mathematical model can be expressed by a recursive formula

$$z_t = b_1 z_{t-1} + b_2 z_{t-2} + \ldots + b_k z_{t-k} + e_t, \qquad (3.129)$$
$$k + 1 \leqq t \leqq T$$

where T is the number of terms in the given realization (the length of the time series observed),
$b_1, b_2, \ldots, b_k$ – regression coefficients related to the correlation function and
e_t – the residual random variable.

Formula 3.129 (often referred to as the linear stochastic regression model) is often used for normally distributed standardized stochastic variables[1]) which facilitates calculation and testing of regression coefficients b_i $(i = 1, 2, \ldots, k)$ and the residual random variable e_t. They are derived by the method of least squares to minimize the variance of the residual random variable e_t. Regression coefficients can be expressed as follows:

$$b_1 = \frac{\mathbf{D}_1}{\mathbf{D}}, \; b_2 = \frac{\mathbf{D}_2}{\mathbf{D}}, \; \ldots, \; b_k = \frac{\mathbf{D}_k}{\mathbf{D}} \qquad (3.130)$$

where $\mathbf{D}$ is the determinant of the set of linear equations, including the correlation coefficients

$$\mathbf{D} = \begin{vmatrix} 1, & r_1, & r_2, & \ldots, & r_{k-1} \\ r_1, & 1, & r_1, & \ldots, & r_{k-2} \\ \vdots & & & & \\ r_{k-1}, & r_{k-2}, & \ldots, & & 1 \end{vmatrix} \qquad (3.131)$$

$\mathbf{D}_i$ is the determinant obtained by replacing the i-th column in $\mathbf{D}$ by the column vector $(r_1, r_2, \ldots, r_k)$.

The standardized residual variance of the random variable e_t is defined as the difference between the total variance and its theoretically "explained" component i.e.

$$s^2(e_t) = m_1(e_t^2) = 1 - b_1 r_1 - b_2 r_2 - \ldots - b_k r_k \qquad (3.132)$$

Often the residual variance (3.132) is calculated for a number of alternatives until its minimum is reached. This alternative also determines the optimal number of terms in the linear expression (3.129) which is equal to the order k of the Markov chain.

---

[1]) Transformation $y = \lg (x - x_0)$ where x is the observed variable and x_0 is the lower boundary, often normalizes the probability distribution (there are also normalizing transformations for the Pearson probability distribution, type III – Kos, 1969; Votruba and Nacházel, 1971). Standardization is performed by means of the formula $z = (y - \bar{y})/s_y$ where $\bar{y}$ is the sample mean and s_y is the sample standard deviation.

Because of the limited length of the observed time series the degrees of freedom are accounted for by

$$\bar{s}^2 = \frac{m_1(e_t^2)\, T}{T - k - 1} \tag{3.133}$$

The values of random variable e_t are generated using the formula

$$e_t = \bar{s} d_t \tag{3.134}$$

where $\bar{s}$ is the standard deviation, i.e. the square root of (3.133), and d_t is the standard random variable (white noise), which can be taken from the tables of random numbers but more often is generated by computers using algorithms for the generation of standard random variables with a given probability distribution.

Having determined the regression coefficients (3.130) and the values of random variable e_t (3.134) we can generate a random sequence of desired length. If standard and transformed variables have been used, an inverse transformation of the sequence z_t is necessary.

In Soviet water resource literature (Svanidze, 1964; Reznikovsky, 1969) special linear regression models are presented, based on the assumption that the conditional c.d.f. and the unconditional c.d.f. belong to the same type and that they can be approximated by the Pearson probability distribution, type III, with $C_s = 2C_v$ (the coefficient of skewness is twice the coefficient of variation).

Stochastic sequences are generated for any order of the Markov chain by the general formula

$$Q_i = \bar{Q}_i - \sum_{j=1}^{\tau} (Q_{i-j} - \bar{Q}_{i-j}) \frac{\sigma_i}{\sigma_{i-j}} \frac{\mathbf{D}_{i(i-j)}}{\mathbf{D}_{ii}} + \Phi_i \sigma_i \sqrt{\frac{\mathbf{D}}{\mathbf{D}_{ii}}} \tag{3.135}$$

where $\mathbf{D}$ is the determinant of the correlation matrix

$$\mathbf{D} = \begin{vmatrix} r_{ii}, & r_{i(i-1)}, & r_{i(i-2)}, & \cdots, & r_{i(i-\tau)} \\ r_{i(i-1)}, & r_{ii}, & r_{i(i-1)}, & \cdots, & r_{i(i-\tau-1)} \\ \vdots & & & & \\ r_{i(i-\tau)}, & \cdots, & & & r_{ii} \end{vmatrix} \tag{3.136}$$

and $\mathbf{D}_{ii}$ and $\mathbf{D}_{i(i-j)}$ are supplements of the coefficients of correlation r_{ii} and $r_{i(i-j)}$ respectively, in the determinant $\mathbf{D}$. Φ_i are the values of standardized random variable with the Pearson c.d.f. with $C_v = 1$ and $C_s = 2C_v = 2$.

For a stationary sequence[1]),

$$\bar{Q}_i = \bar{Q}_{i-1} = \ldots = \bar{Q}_{i-j} = \bar{Q}_{i-\tau} = \bar{Q}$$

$$\sigma_i = \sigma_{i-1} = \ldots = \sigma_{i-j} = \sigma_{i-\tau} = \sigma$$

and the equation can be simplified as

$$Q_i = \bar{Q} - \sum_{j=1}^{\tau} (Q_{i-j} - \bar{Q}) \frac{\mathbf{D}_{i(i-j)}}{\mathbf{D}_{ii}} + \Phi_i \sigma \sqrt{\frac{\mathbf{D}}{\mathbf{D}_{ii}}} \tag{3.137}$$

The conditional standard deviation and the coefficient of skewness are determined by the expressions

$$\sigma_{\text{cond}} = \sigma \sqrt{\frac{\mathbf{D}}{\mathbf{D}_{ii}}} \tag{3.138}$$

$$C^i_{\text{s,cond}} = \frac{2\sigma \sqrt{\frac{\mathbf{D}}{\mathbf{D}_{ii}}}}{\bar{Q} - \sum_{j=1}^{\tau} (Q_{i-j} - \bar{Q}) \frac{\mathbf{D}_{i(i-j)}}{\mathbf{D}_{ii}}} \tag{3.139}$$

If the stochastic sequence is calculated in relative values then (3.137) becomes

$$k_i = 1 - \sum_{j=1}^{\tau} (k_{i-j} - 1) \frac{\mathbf{D}_{i(i-j)}}{\mathbf{D}_{ii}} + \Phi_i C_v \sqrt{\frac{\mathbf{D}}{\mathbf{D}_{ii}}} \tag{3.140}$$

3.2.6.2 Model of Generation of Monthly Stochastic Variables

Generation of stochastic hydrological time series with a time interval shorter than one year is more difficult than generation of annual time series, because the model has to take into account the fluctuations of probability distributions of random variables during the year.

---

[1]) Stochastical hydrological time series have been generated for stationary processes and some attempts have been made to use non-stationary processes as well. In the Hydrological Department of the Faculty of Civil Engineering (Technical University of Prague) research into some non-stationary hydrological time series has been carried out on the basis of differences, in development of particular statistical characteristics (Nacházel *et al.* 1975).

Annual flows were generated on the basis of periodic models. Their deterministic component is given (as compared with the regression models) by a harmonic function, the form of which is optimized by spectral analysis. The random component has the same meaning as in the regression models, and it is often generated as white noise (Anděl and Balek, 1969; Balek, 1975).

There are some special types of mathematical models formed as the sum of the deterministic periodic and random regression components.

Svanidze's disaggregation method (1964) used the technique of double sampling and provided a relatively good fit between input and output annual statistical parameters. However, disaggregation using the monthly "fragments", in combination with the generated annual flows, distorted some statistical characteristics of monthly flows, and application of the method without careful analysis could result in incorrect results (Votruba and Broža, 1966).

Remarkable progress in the generation of monthly flows was achieved by Kos, 1969. From many samples of observed time series, he proved that the principle of the linear regression model can be applied to monthly flows where all possible combinations of correlation relationships are involved (e.g. between the time series of monthly flows in February and January, March and January, etc.). Expression (3.141) constructed analogously to (3.129) is the regression equation for the generation of sequences of monthly values in a chosen month m

$$z_{c,m} = b_{1,m}z_{c,m-1} + b_{2,m}z_{c,m-2} + \ldots + b_{k,m}z_{c,m-k} + e_m \tag{3.141}$$

where m is 1, 2, ..., 12 (month of the year)

c – 1, 2, ..., $T/12$ (number of generation cycle – i.e. year, where T = the total number of month generated)

e_m – random deviate in month m

The monthly generating model requires twelve equations of type (3.141). The generating process uses the step equal to 12, i.e. in annual cycles: having chosen 12 random variables in the first year (e.g. by observed values or by monthly means), in the second cycle the values $z_{2,m}$ are calculated etc. In this manner a sequence of random variables in chronological order is calculated analogously to the observed time series. The unknown regression coefficients $b_{i,m}$ and random deviates e_m are determined in the same way as in the annual model. The regression coefficients are determined by

$$b_{1,m} = \frac{\mathbf{D}_{1,m}}{\mathbf{D}_m}, \; b_{2,m} = \frac{\mathbf{D}_{2,m}}{\mathbf{D}_m}, \; \ldots, \; b_{k,m} = \frac{\mathbf{D}_{k,m}}{\mathbf{D}_m} \tag{3.142}$$

where the determinants $\mathbf{D}_m$ and $\mathbf{D}_{i,m}$ $(i = 1, 2, \ldots, k)$ have the same meaning as in the annual model but are calculated for each month m.

The residual variance in each month m is calculated by the equation

$$s^2(e_m) = m_1(e_m^2) = 1 - b_{1,m}r_{1,m} - b_{2,m}r_{2,m} - \ldots - b_{k,m}r_{k,m} \tag{3.143}$$

A number of alternatives are used in order to minimize this variance and the corresponding k is the order of the Markov chain.

Because of the limited length of the series

$$\bar{s}_m^2 = \frac{m_1(e_m^2)\, C}{C - k - 1} \tag{3.144}$$

where

$$C = \frac{T}{12}.$$

The random variates e_m in each month are calculated by

$$e_m = \bar{s}_m d_t \tag{3.145}$$

The generating procedure has already been described. The last stage of calculation is the inverse transformation of the generated values if standardization has been used.

The goodness-of-fit of the linear regression model has been tested many times in water resource research and hydrological studies by statistical and other tests. The goodness-of-fit of input and output parameters was established and the deviations were not statistically significant. If the order k is less than 12 the annual parameters are not taken into account. A possible improvement of the model is being tested (Nacházel *et al.*, 1975; Kos, 1982).

In future development of monthly hydrological time series generation, the periodic models could be tested according to the obvious annual cycle.

Attempts to use the methods described for daily hydrological time series have failed; further research is desirable in this respect.

3.2.6.3 Model of Flow Generation in a System of Stations

The WRS model often requires as input values the monthly flows generated in a system of stations. In this system the elements are stochastically dependent. Therefore the stochastic time series has to be generated in a system to maintain the cross-correlations, i.e. the cross-correlations of the generated time series should correspond to those of the observed time series. The model of a system of stations should involve two subsystems of correlation relations, viz.:

(1) The inner serial correlation of individual stations (with the length given by the Markov chain) and

(2) cross-correlations among the flows in different stations.

The simplest method of modelling a system of stations is that of the central and satellite stations. Using the linear regression model (see section 3.2.6.2) the flows are generated in the central (the most important) station. Coincident flows in other satellite stations are calculated by simple linear regression

$$y_{i,j,k} = \bar{y}_{j,k} + b_{j,k}(x_{i,j} - \bar{x}_j) + u_i d_{j,k} \tag{3.146}$$

$$d_{j,k} = s_{j,k}\sqrt{1 - r_{j,k}^2} \tag{3.147}$$

where $y_{i,j,k}$ is the flow of serial number i, in month j, at point k,
$\bar{y}_{j,k}$ – mean monthly flow in month j and at point k,
$b_{j,k}$ – regression coefficient in month i for the relationship between the central station and the satellite station k,
$x_{i,j}$ – flow in the central station of serial number i in month j,
$\bar{x}_j$ – mean monthly flow in the central station in month j,
u_i – standardized random variate,
$s_{j,k}$ – standard deviation in month j at point k
$r_{j,k}$ – correlation coefficient in month j for relationship of flows between the central station and the satellite station k.

Calculation by (3.146) and (3.147) proceeds month after month. For each combination of the central station with a satellite station 12 equations (3.146) and (3.147) are necessary.

This method can be applied to synchronous or almost synchronous time patterns. If the cross-correlation is low, this method may distort the serial correlation in the satellite station.

Orthogonal transformation is a more general method of generating flows in a system of stations. The principle here is involved in transformation of linearly dependent (coincident) flows in each month in different stations to independent (orthogonal) values. It is assumed that the observed flows $x_{i,j,k}$ in the same month j but in different stations k form correlated random vectors that are transformed to independent vectors by the expression

$$\mathbf{z}_{i,j,k} = q_{j,1,k}x_{i,j,1} + q_{j,2,k}x_{i,j,2} + \ldots + q_{j,n,k}x_{i,j,n} \tag{3.148}$$

where $i = 1, 2, \ldots, T$, (number of year of observation),
$j = 1, 2, \ldots, 12$ (number of month),
$k = 1, 2, \ldots, n$ (number of station)

In each month j the coefficients $q_{j,n,k}$ are functions of the correlation matrix with the elements given by correlation coefficients between the coincident values $x_{i,j,k}$ for different k values. There are twelve such square matrices with the dimension $n \cdot n$ (according to the number of stations), each in the form

$$\begin{matrix} r_{j,1,1}, & r_{j,1,2}, & \ldots, & r_{j,1,n} \\ r_{j,2,1}, & r_{j,2,2}, & \ldots, & r_{j,2,n} \\ \vdots & & & \\ r_{j,n,1}, & r_{j,n,2}, & \ldots, & r_{j,n,n} \end{matrix} \quad \text{where } j = 1,2, \ldots, 12 \tag{3.149}$$

The matrix (3.149) is symmetrical about the principle diagonal, all the elements of which are 1 (correlation of the flows in the same month and station).

The orthogonal transformation produces independent flows in fictitious stations.

Its advantage is that the flows in each station can be generated independent of other stations and for any conditions including the order of the Markov chain. After transformation of the observed flows in all the stations individual time series are generated for individual stations. Having finished this calculation, we proceed to the inverse transformation that produces the output correlated stochastic flows in real stations. The inverse transformation is described by the following equation

$$x_{i,j,k} = p_{j,1,k}z_{i,j,1} + p_{j,2,k}z_{i,j,2} + \ldots + p_{j,n,k}z_{i,j,n} \tag{3.150}$$

The coefficients $p_{j,n,k}$ are the elements of the inverse matrix of the original matrix composed of elements $q_{j,n,k}$. Because this is an orthogonal matrix, its inverse matrix is the transposed matrix where the rows are replaced by columns and vice versa.

Calculation of the coefficients $q_{j,n,k}$ or $p_{j,n,k}$ is relatively complicated, particularly for a large number of stations. Therefore, computer software including special algorithms (e.g. Jacobi procedure) can be found in specialized mathematical literature (Evans, 1962; Naur, 1963).

3.3 MATHEMATICAL DESCRIPTION OF SYSTEM BEHAVIOUR

3.3.1 Introduction

Assume a system with all inputs and outputs specified and described. An analysis of the system's behaviour involves investigation of interrelations between all these inputs and outputs and, in addition, the relationships among all the n elements of the system and analysis of the individual behaviour of each element. Such a complex analysis may be difficult even for a relatively limited system.

The *behaviour of the whole system* is given by the interrelations between its inputs and outputs. The input X is given by p input components ($p \geqq 1$) $x_1, x_2, \ldots, x_p$ so that it can be described by the input vector

$$\bar{x} = (x_1, x_2, \ldots, x_p) \tag{3.151}$$

The output Y is given by q output components ($q \geqq 1$) and it can be similarly described by the output vector

$$\bar{y} = (y_1, y_2, \ldots, y_q) \tag{3.152}$$

The behaviour can be symbolically described by transformation $\bar{T}$

$$\bar{y} = \bar{T}(\bar{x}) \tag{3.153}$$

$\bar{T}$ is often called the transformation operator.

This behaviour can be *deterministic* or *stochastic* – the latter being the general case of system behaviour. Deterministic behaviour can be classified as combinatory of sequential. In *combinatory* behaviour, the transformation $\bar{T}$ is unique, i.e. only

one output vector $\bar{y}$ corresponds to a certain vector $\bar{x}$. In *sequential behaviour*, different responses correspond to the same stimulus in relation to the previous states defined or to their sequences.

In this section we shall investigate that class of *dynamic systems* which have time-variant general properties.

3.3.2 The Concept of Dynamic System Operator

A dynamic system (Fig. 3.19) with the input time-dependent function (process) $X(t)$ and output process $Y(t)$ is described by

$$Y(t) = A\{X(t)\} \tag{3.154}$$

where A is the *dynamic system operator*. It symbolizes the set of all mathematical and logical operations applied inside the system S on the input process $X(t)$ in such a way that it is transformed into the output process $Y(t)$.

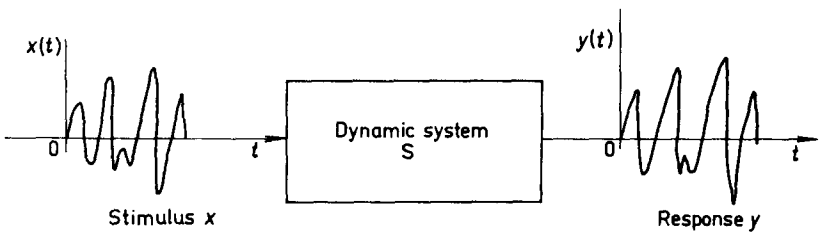


Fig. 3.19 Schematic representation of the dynamic system

The concept of the operator is a generalization of the concept of the functional (see section 3.7), just as the functional is a generalization of the concept of a function. In the case of the operator, a function (process) as the "independent variable" is related again to a function (process) as the "dependent variable".

The operator can be either linear or non-linear. *Linear operators* describing the transformation of the inputs of a linear dynamic system are characterized by satisfying the superposition principle, i.e. for any input functions $x_1(t)$, $x_2(t)$, ..., $x_n(t)$ and for any constant n, c_1, c_2, ..., c_n,

$$A\left\{\sum_{k=1}^{n} c_k\, x_k(t)\right\} = \sum_{k=1}^{n} c_k\, A\{x_k(t)\} \tag{3.155}$$

The result of the realization of the linear operator A to any linear combination of input functions is equivalent to the linear combination of the results of realization of this operator to the individual input functions.

Similarly, for an input function given by a set of differential elementary stimuli, the output function of a linear dynamic system is given by the integral of the corresponding differential responses i.e.

$$A\left\{\int_{\lambda_1}^{\lambda_2} c(\lambda)\, x(t, \lambda)\, \mathrm{d}\lambda\right\} = \int_{\lambda_1}^{\lambda_2} c(\lambda)\, A\{x(t, \lambda)\}\, \mathrm{d}\lambda \tag{3.156}$$

A distinction can be draw between *homogeneous* (e.g. linear operator of differentiation, integration, multiplication by a function of time, etc.) and *non-homogeneous* linear operators, which are composed of a homogeneous operator and a non-random function of time.

The behaviour of the dynamic system S can be described by a linear differential equation with constant coefficients

$$a_n \frac{\mathrm{d}^n y(t)}{\mathrm{d}t^n} + a_{n-1} \frac{\mathrm{d}^{n-1} y(t)}{\mathrm{d}t^{n-1}} + \ldots + a_1 \frac{\mathrm{d}y(t)}{\mathrm{d}t} + a_0\, y(t) =$$

$$= b_m \frac{\mathrm{d}^m x(t)}{\mathrm{d}t^m} + b_{m-1} \frac{\mathrm{d}^{m-1} x(t)}{\mathrm{d}t^{m-1}} + \ldots + b_1 \frac{\mathrm{d}x(t)}{\mathrm{d}t} + b_0\, x(t) \tag{3.157}$$

or by a symbolic linear differentiation operator $\mathrm{d}^k/\mathrm{d}t^k = p^k$ in the form

$$(a_n p^n + a_{n-1} p^{n-1} + \ldots + a_1 p + a_0)\, y(t) =$$

$$= (b_m p^m + b_{m-1} p^{m-1} + \ldots + b_1 p + b_0)\, x(t) \tag{3.158}$$

or, more briefly,

$$A_n(p)\, y(t) = B_m(p)\, x(t) \tag{3.159}$$

Hence

$$y(t) = \frac{B_m(p)}{A_n(p)}\, x(t) \tag{3.160}$$

The ratio $B_m(p)/A_n(p)$ is an important characteristic of the operator of a linear dynamic system, and it is called the *transfer of the system.*

The *non-linear operator* (operator of the second and higher powers of the input function, operator of trigonometric functions, etc.) similarly describes non-linear dynamic systems.

Example VII

Assume a system S with input process $x(t)$ and output process $y(t)$.

Determine its operator if

$x(t) = a \cos(t)$ and $y(t) = -a \sin(t)$

Solution

Denote the operator of this system by A. Since

$$\frac{\mathrm{d}(a \cos t)}{\mathrm{d}t} = -a \sin t \tag{VII.1}$$

the operators A equals

$$A = \frac{\mathrm{d}}{\mathrm{d}t} \tag{VII.2}$$

A is the operator of the system S and it is called the differentiation operator. This operator is linear and homogeneous if relationships (3.155) and/or (3.156) are satisfied.

3.3.3 Dynamic System with Random Inputs and Outputs

The input and output functions of a system can be random time functions. If these functions are stationary, i.e. invariant to the time origin, the system is called a *stationary dynamic system*; the opposite is a *non-stationary dynamic system* where a change in time of occurrence of the same stimulus results in a change in the response (Fig. 3.20).

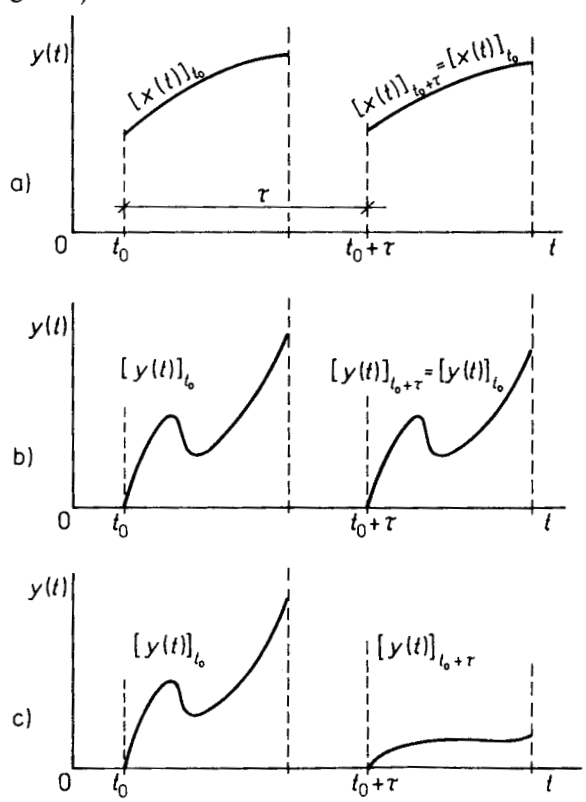


Fig. 3.20. Schematic representation of the response of stationary and non-stationary systems
a – stimulus, *b* – response of a stationary system, *c* – response of a non-stationary system

Random process occur frequently in the inputs and outputs of dynamic systems in WRS contexts. In general, it is assumed that the input and output processes are non-stationary. Assume a non-stationary dynamic system described by the linear operator A_L. The homogeneous linear operator A_0 can be separated in the transformation, i.e.

$$Y(t) = A_L\{X(t)\} = A_0\{X(t)\} + \varphi(t) \tag{3.161}$$

where $\varphi(t)$ is a non-random time function.

An input stochastic process $X(t)$ can be described, in general, by a non-stationary mean $m_x(t)$, standard deviation $s_x(t)$, autocorrelation function $r_x(t, \tau)$ and possibly by an adequate theoretical probability distribution $f_x(x, t)$. Similarly an output stochastic process $Y(t)$ is described by quantities $m_y(t)$ and $s_y(t)$ and functions $r_y(t, \tau)$ and $f_y(y, t)$, which can be derived from the input data. Then we get, for example, for $m_y(t)$

$$m_y(t) = M[A_0\{x(t)\}] + \varphi(t) \tag{3.162}$$

where A_0 is the homogeneous linear operator and M [ ] means mean value (expectation) of the expression in brackets.

The calculus of operators was developed for the solution of differential equations. It is closely related to Laplace integral transformations. The symbols of the calculus of operators can be applied in problems of input/output relationships of dynamic systems.

Sometimes it is useful to represent mathematically the relationship between the inputs and outputs by the convolution integral in the form

$$y(t) = \int_0^t x(\tau)\, h(t - \tau)\, d\tau \tag{3.163}$$

where $x(\tau)$ is the input time function,
$y(\tau)$ – the output time function,
$h(t - \tau)$ – the transfer function,
t, τ – time arguments,
$(t - \tau)$ – the time lag (difference in time between the occurrence of stimulus and response).

The transfer function $h(t - \tau)$ is a characteristic of the transformation operator of dynamic systems, and it is similar to the transfer of linear dynamic systems. A further generalization the concept is the notion of the kernel of the system (see section 3.3.4).

3.3.4 Mathematical Formulation of the Behaviour of Stochastic Hydrological Systems

Dynamic systems with stochastic behaviour are frequently encountered in WRS investigation. Moreover, these stochastic systems are, in general, non-stationary and their behaviour is described by non-linear transformations.

The general form of the above-mentioned mathematical description is particularly applicable to some hydrological stochastic systems.The transformation of inputs to outputs can be given in the following form, assuming the linearity of the model (Strupczewski, 1975).

$$y(t) = H_d[x_d(t)] + H_s[x_s(t)] \tag{3.164}$$

where $x_d(t)$ is the deterministic input component (deterministic input signal),
$x_s(t)$ – stochastic input component (random input signal),
H_d – linear operator of the deterministic input signal,
H_s – linear operator of the stochastic input signal.

This concept is related to some hydrological stochastic models and their transformations (Kisiel, 1969, Yevjevich, 1976, Dooge, 1972). The behaviour of a compound linear hydrological system can be described by a convolution integral in the following general form (Huthmann, 1975)

$$y(t) = \sum_{i=1}^{N} \int_0^{\infty} h_i(\tau)\, x_i(t - \tau)\, d\tau + z(t) \tag{3.165}$$

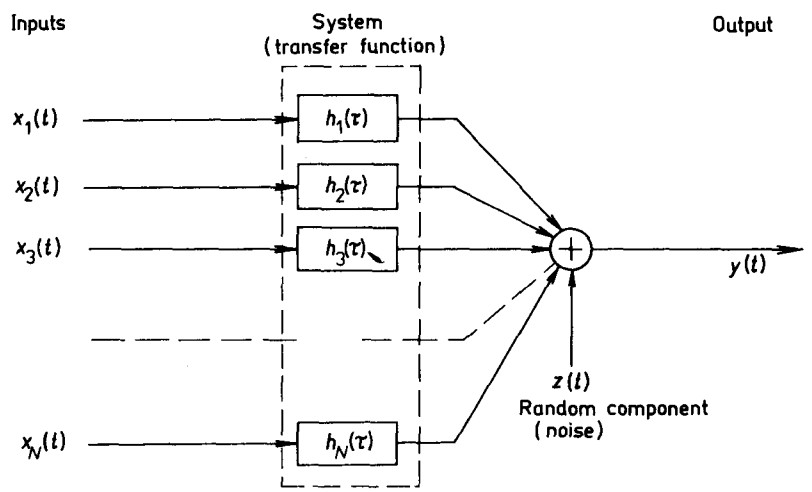


Fig. 3.21 Schematic representation of a compound linear stochastic system

where N is the number of inputs,
$y(t)$ – output,
$x_i(t)$ – i-th input,
$h_i(\tau)$ – i-th transfer function related to the i-th input,
$z(t)$ – random component (noise).

A schematic representation of this general form is given in Fig. 3.21. This formal relationship was applied by G. Huthmann in the determination of the maximum basin runoff.

A similar, more complex case with a non-linear transformation operator is given in general form in equation (3.166). The behaviour of the system is described as a multiple-dimensional convolution integral

$$y(t) = \sum_{j=1}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_j(\tau_1, \tau_2, \dots, \tau_j) \prod_{i=1}^{n} x(t' - \tau_i)\, d\tau_1, \dots, d\tau_j \tag{3.166}$$

where $x(t)$ is the system input,
$y(t)$ – the system output,
$h_j(\tau_1, \tau_2, \dots, \tau_j)$ – is the kernel of the system[1])

This approach to the description of a complex, non-linear system involving the determination of the kernels of several orders was used by N. Wiener. In hydrology contexts, only the kernel of the first order, i.e. the transfer function, has a clear physical meaning. The kernels of higher orders have to be interpreted. Quimpo modified the relationship (3.166) to a finite sum of integrals in the form

$$y(t) = \sum_{j=1}^{m} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_j(\tau_1, \tau_2, \dots, \tau_j)\; x(t - \tau_1)\, x(t - \tau_2) \dots x(t - \tau_j)\, d\tau_1\, d\tau_2 \dots d\tau_j \tag{3.167}$$

and found the solution of the kernel of the second order $h_2(\tau_1, \tau_2)$ which is a certain generalization of the transfer function.

The non-linear model of the rainfall-runoff relationship, the behaviour of which is described by the general equation (3.167) was successfully used for the n-day forecast of flows. This application showed the possibilities and advantages of this approach to the treatment of problems in stochastic hydrological systems. However, further applications are envisaged after comprehensive improvement of mathematical procedures aimed at discretization of the relationships describing the behaviour of the system and after a deeper physical interpretation of the kernels of the system, which are generalizations of the transfer function concept.

[1]) The concept of the kernel is defined in the solution of integral equations (Rektorys, 1963). The kernel is a function $K(x, s)$ in an integral equation of the second kind in the form

$$f(x) - \int_a^b K(x, s)\, f(s)\, ds = g(x)$$

3.4 FOURIER TRANSFORMATION

3.4.1 Introduction

In the investigation of WRS hydrological processes the concept of spectrum and the spectral density function (s.d.f.) are often used (Votruba and Nacházel, 1969). In the theory of spectral analysis, a transformation from the linear time argument to the frequency characteristic is performed, which facilitates and improves the investigation of complicated stochastic and deterministic processes. Various transformations are necessary in the determination of relationships between WRS inputs and outputs. The choice and analysis of such transformations in various present-day tasks of water management by means of the systems approach has not yet been developed to the stage of implementation so their practical importance is not obvious. However, they are promising elements of the theoretical basis. In inputs and outputs of dynamic systems, processes (time functions) of various types can be found. The basic classification criterion is their *deterministic* or *stochastic nature*. Deterministic (non-random) processes can be described by exact mathematical relationships, stochastic (random) processes cannot. The deterministic processes can be further subdivided into several groups (Fig. 3.22).

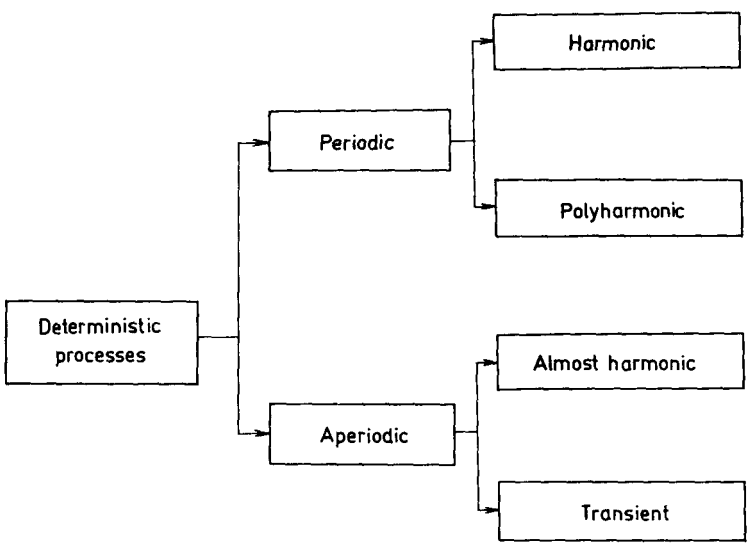


Fig. 3.22 Deterministic processes classification

Harmonic processes of a periodic function can be described in a compact and simple form by an elementary discrete spectrum with one non-zero value (see section 3.2.4 and Fig. 3.14a).

A polyharmonic process, which occurs more often in reality, is formed by the combination of two of more elementary harmonic processes, assuming that the ratio of the lengths of the periods for each pair of harmonic processes are rational numbers. In the opposite case the process in question is called almost periodic; it is composed of sine curves not only of divisible periods but with all possible periods where also an infinite period may appear. These cases occur if two or more independent harmonic processes are combined. For the polyharmonic or almost harmonic processes the spectrum can be expresses by several discrete values; in the case of almost harmonic processes some remainders are left over.

The class of transient processes includes all the other cases of periodic processes that occur in many physical contexts. The difference between harmonic, polyharmonic and almost harmonic processes, on the one hand, and transient processes, on the other, is obvious from the different form of their spectra. Transient processes cannot be expressed by a discrete spectrum. In almost all cases, however, their spectra can be described by a continuous function using the Fourier integral, which facilitates the transformation of all these types of processes to a comparable level.

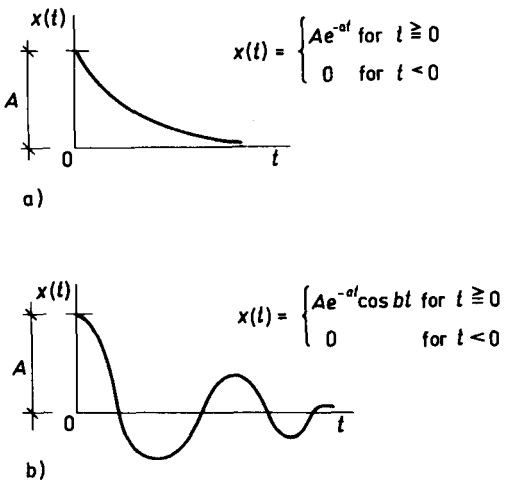


Fig. 3.23 Examples of aperiodic processes and their spectra

Similar relationships hold for stochastic processes and therefore the use of the Fourier integral and the Fourier transformation facilitates the application of the unique method of investigation of processes occurring in the inputs and outputs of dynamic systems and of their relationships irrespective of the different characteristics of these processes.

Figure 3.23 shows an example of two transient processes with continuous spectra. Both these time functions can be applied in hydrology and WRS, the first case (a) being an exponential function and the second case (b) a function approximately har-

monically damped. Changes taking place in them during the transition of the process through the system will be followed up. Using the argument of the spectral function, which is $\omega = 1/T$ (where T is analogous to the length of the period of periodic functions), the Fourier integral may be expressed in the form:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-2\pi i\omega t}\, dt \tag{3.168}$$

where i is the imaginary unit $\sqrt{-1}$ (in the general case, spectrum is a complex function) and $x(t)$ is the input process with the time argument t, the spectrum of which is to be found.

Using the modulus (mean value) of the spectrum $|X(\omega)|$ and the argument $\Theta(\omega)$, the complex form may be written

$$|X(\omega)| = X(\omega) \cdot e^{i\Theta(\omega)} \tag{3.169}$$

The spectrum of function (a) has a broad-band character (Fig. 3.14d), whereas that of function (b) has a narrow-band character (Fig. 3.14b).

3.4.2 Fourier Series

The periodic function $x(t)$ with period T can be represented as the sum of a finite or infinite number of simple harmonic components (sine curves). If the periodic function analysed, $x(t)$, is bounded, piecewise continuous, and if it has inside the period T a finite number of extrema, it can be written in the form

$$x(t) = A_0 + \sum_{k=1}^{\infty} A_k \sin(k\omega t + \alpha_k) \tag{3.170}$$

where $A_0, A_1, \ldots, A_k$; $\alpha_1, \alpha_2, \ldots, \alpha_k$ are constants corresponding to the particular harmonic components and ω, the frequency, equals $2\pi/T$.

This equation is the general expansion for a periodic function $x(t)$ into a trigonometric series. If a new time argument is introduced

$$\tau = \omega t = \frac{2\pi t}{T} \tag{3.171}$$

so that

$$t = \frac{\tau}{\omega} \tag{3.172}$$

its function $f(\tau)$ can be written in the form

$$f(\tau) = a_0 + \sum_{k=1}^{\infty} \left(a_k \cos 2\pi k \frac{t}{T} + b_k \sin 2\pi k \frac{t}{T} \right) \tag{3.173}$$

where $a_0 = A_0$,
$a_k = A_k \sin \alpha_k$,
$b_k = A_k \cos \alpha_k$,
$(k = 1, 2, 3, \ldots)$.

The expansion of the function $x(t)$ and $f(\tau)$ in the form (3.173) is called the Fourier series in the real form. In complex (exponential) form it is

$$f(\tau) \approx \sum_{k=-\infty}^{\infty} c_k \, e^{ik\tau} \tag{3.174}$$

where

$$c_0 = \frac{a_0}{2},$$

$$c_k = \frac{a_k - ib_k}{2} \quad \text{if} \quad k > 0,$$

$$c_k = \frac{a_k + ib_k}{2} \quad \text{if} \quad k < 0, \tag{3.175}$$

i is the imaginary unit
or in a complex representation by one equation for c_k

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\tau) \, e^{-ik\tau} \, d\tau \tag{3.176}$$

where $k = 0, \pm 1, \pm 2, \ldots$.

The expansion of a periodic function into a Fourier series simplifies the mathematical expression of very complicated functions that are important for technical applications. The known spectrum of a periodic function facilitates analysis of the behaviour of a system if the periodic functions are applied in the system input. This holds for both deterministic and stochastic periodic functions, assuming that the spectrum can be represented in this form.

3.4.3 Fourier Integral

The Fourier integral is a generalization of the expansion of functions, including aperiodic functions. In the Fourier series expansion the number of harmonic functions could be infinite, but the number of the frequencies was finite. If the sum of an infinite number of simple harmonic functions is expressed with frequencies that differ by an increment approaching zero, a Fourier integral results. However, it exists and it can be calculated under certain conditions called *conditions of convergence.*

Figure 3.24 shows a schematic representation of the spectrum of an aperiodic function given by a continuous or piecewise continuous curve, and the discrete (point) spectrum of a polyharmonic function with three components T_1, T_2, T_3 with frequencies ω_1, ω_2, ω_3, respectively. The spectrum of an aperiodic function can be described by a continuous, often bell-shaped curve, with the ranges of frequencies (ω_a, ω_b) and differences between successive frequencies equal to the differential d_ω.

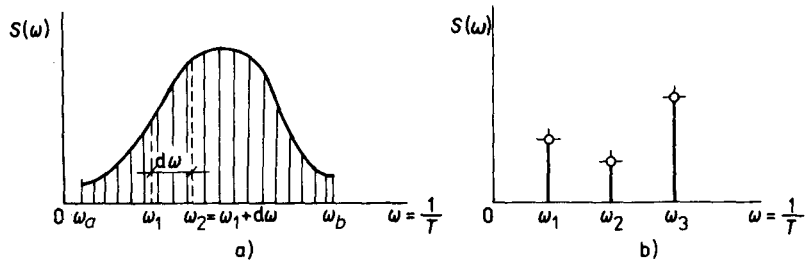


Fig. 3.24 Schematic representation of spectra of aperiodic and periodic functions

A sufficient condition of the Fourier integral convergence for the function $f(\tau)$ is the existence of the improper integral

$$I_n = \int_{-\infty}^{\infty} |f(\tau)| \, d\tau \tag{3.177}$$

the function $f(\tau)$ is assumed to be piecewise continuous with a finite number of discontinuities and local extrema in each finite interval $\Delta\tau$. At each point τ_0 inside the finite interval $(-\lambda; \lambda)$ (and $|\lambda| > \tau_0$; the function $f(\tau)$ is differentiable),

$$f(\tau_0) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos \frac{k\pi\tau_0}{\lambda} + b_k \sin \frac{k\pi\tau_0}{\lambda} \right) \tag{3.178}$$

If the coefficients a_0, a_k, b_k of this expansion are expressed in an integral form, we get[1]):

$$f(\tau_0) = \frac{1}{\pi} \int_0^{\infty} d\omega \int_{-\infty}^{\infty} f(\tau) \cos \omega (\tau - \tau_0) \, d\tau \tag{3.179}$$

For formal reasons we replace τ_0 by τ and τ by t:

$$f(\tau) = \frac{1}{\pi} \int_0^{\infty} d\omega \int_{-\infty}^{\infty} f(t) \cos \omega (t - \tau) \, dt \tag{3.180}$$

[1]) A detailed mathematical description can be found in the mathematical literature (Lapa, 1971; Raven, 1966).

If the function $f(t)$ is integrable in the interval $(-\infty; \infty)$, and if it possesses in any finite interval Δt a finite number of discontinuities, then the expression (3.180) holds for every point where it can be differentiated, and it is the Fourier integral. This integral (rather, two-dimensional integral) can be written in a complex form

$$f(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} f(t)\, e^{i\omega(\tau - t)}\, d\tau \tag{3.181}$$

which is related to the complex (exponential) form of the Fourier series.

3.4.4 Direct and Inverse Fourier Transformation

The complex form of the Fourier integral can be modified to the form

$$f(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(t)\, e^{-i\omega t}\, dt \right] e^{i\omega\tau}\, d\omega \tag{3.182}$$

where

$$\int_{-\infty}^{\infty} f(t)\, e^{-i\omega t}\, dt = F(\omega) \tag{3.183}$$

This relationship between the complex s.d.f., $F(\omega)$, and the real function of time $f(\tau)$ is called the *Fourier transformation* (direct Fourier transformation) (Fig. 3.25), and for the time argument defined by (3.171) is has the following form

$$F(\omega) = \int_{-\infty}^{\infty} f(\tau)\, e^{-i\omega t}\, d\tau \tag{3.184}$$

Expressing the function $f(\tau)$ by its s.d.f. significantly facilitates the investigation of dynamic system behaviour as a response to this function.

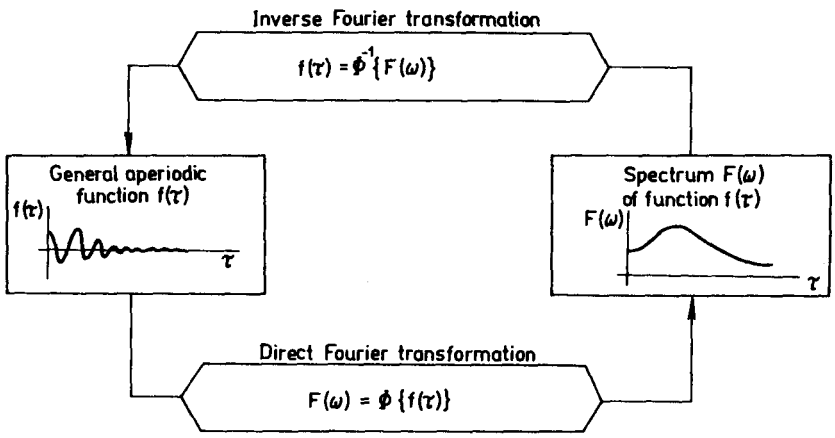


Fig. 3.25 Schematic representation of the direct and inverse Fourier transformation

The reverse of the direct Fourier transformation in the form (3.184) is the *inverse Fourier transformation* in the form

$$f(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)\, e^{i\omega\tau}\, d\omega \tag{3.185}$$

Instead of the complex form (3.185) the real form of the inverse Fourier transformation can be used

$$f(\tau) = \frac{1}{\pi} \int_{0}^{\infty} [A(\omega) \cos \omega\tau - B(\omega) \sin \omega\tau]\, d\omega \tag{3.186}$$

where

$$A(\omega) + i\, B(\omega) = F(\omega) \tag{3.187}$$

where $A(\omega)$ is an event function and $B(\omega)$ is an odd function.

3.4.5 Spectrum of a Unit Impulse

In the theory of dynamic systems and their models the unit impulse is often applied to their inputs. It can be described by the Dirac delta function (Dirac unit impulse)

$$f(t) = 0 \quad \text{for} \quad \tau \in \left(-\infty;\ -\frac{\varepsilon}{2}\right) \tau \in \left(\frac{\varepsilon}{2};\ \infty\right)$$

$$f(\tau) = \frac{1}{\varepsilon} \quad \text{for} \quad \tau \in \left(-\frac{\varepsilon}{2};\ \frac{\varepsilon}{2}\right) \tag{3.188}$$

where ε is a positive number that converges to zero.

The *Dirac function* is aperiodic and if eq. (3.187) is used for it, we get the following Fourier transform

$$F(p) = \int_{-\infty}^{\infty} f(\tau)\, e^{-p\tau}\, d\tau = \int_{-\varepsilon/2}^{\varepsilon/2} \frac{1}{\varepsilon}\, e^{-p\tau}\, d\tau = \frac{1}{\varepsilon_p} \left(e^{\varepsilon_p/2} - e^{-\varepsilon_p/2}\right) \tag{3.189}$$

The limit of (3.189) for the duration of unit impulse $\varepsilon \to 0$ is

$$\lim_{\varepsilon \to 0} \left[\frac{1}{\varepsilon_p} \left(e^{\varepsilon_p/2} - e^{-\varepsilon_p/2}\right)\right] = 1 \tag{3.190}$$

The Dirac unit impulse of the length ε_p approaching zero and its spectrum including all possible frequencies from zero to infinity with a constant amplitude are shown in Fig. 3.26.

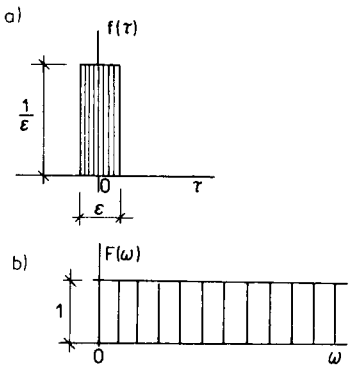


Fig. 3.26 Schematic representation of the Dirac impulse and its spectrum
a – Dirac unit impulse, *b* – spectrum of the Dirac unit impulse

3.4.6 Characteristic Function of Random Variable

The random variable X can be described by its mean value $M[X]$, p.d.f., $f(x)$ and other parameters, but, in addition, it can be handled in some random variable operations, by the mean value of the complex exponential function

$$M[e^{itX}] = g(t) \tag{3.191}$$

which is called the *characteristic function* of this random variable. This representation was introduced by Lyapunov (Ventcel, 1962).

The relationship between the p.d.f. and its characteristic function is, in fact, the direct Fourier transformation, which can be written for a continuous random variable in the form

$$g(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx \tag{3.192}$$

and, for the discrete random variable with probabilities p_k corresponding to random variable values x_k, in the form

$$g(t) = \sum_{k=1}^{n} e^{itx_k} p_k \tag{3.193}$$

The inverse Fourier transform is

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-itx} g(t)\,dt \tag{3.194}$$

The two basic properties of characteristic functions are:

(1) If X and Y are random variables related by the linear transformation $Y = aX$ where a is a non-random coefficient, then

$$g_y(t) = g_x(at) \tag{3.195}$$

(2) Assume n random variables $X_1, X_2, \ldots, X_n$ and their characteristic functions $g_{x_1}(t), g_{x_2}(t), \ldots, g_{x_n}(t)$, then the characteristic function of the random variable

$$Y = \sum_{k=1}^{n} X_k$$

has the following form

$$g_y(t) = g_{x_1}(t)\, g_{x_2}(t) \ldots g_{x_k}(t) \ldots g_{x_n}(t) = \prod_{k=1}^{n} g_{x_k}(t) \tag{3.196}$$

An example from hydrology and WRS problems could be to determine the p.d.f. of a random variable Z which is the sum of two independent random variables X and Y with known p.d.f.s.

Using the second basic property of the characteristic function we get from $g_x(t)$ and $g_y(t)$

$$g_z(t) = g_x(t)\, g_y(t) \tag{3.197}$$

hence

$$f_y(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}tz}\, g_z(t)\, \mathrm{d}t \tag{3.198}$$

However, the same problem can be solved by other methods, e.g. using the convolution principle (see section 3.5.7).

3.5 LAPLACE TRANSFORMATION

3.5.1 Introduction

The Laplace Fourier integral transformations are the basis for the description of dynamic system behaviour. Just as for the Fourier transformation, the principal idea underlying the Laplace transformation is the transformation of the time argument to the frequency one; the investigation of the input real time function is replaced by an analysis of its spectrum, which is generally a complex function. In many applications the conditions of convergence of the Fourier integral transformation (3.181) are not satisfied and its calculation is not possible. In this sense, the Laplace transformation is more useful, particularly for technical problems.

The use of integral transformations in research of systems behaviour in hydrology, theory of storage and WRS is not yet extensive enough to allow an evaluation of the potential advantages of each method; however, they form an important theoretical basis.

The Laplace transformation and its modifications replace aperiodic functions by the sum of an infinite number of damped or exponentially increasing harmonic components, whereas the Fourier transformation applies the simple (not damped) harmonic oscillations. The existence and applicability of the Laplace transformation is not universal; for the condition of its convergence see section 3.5.2.

3.5.2 Direct and Inverse Laplace Transformation

The direct Laplace transformation (or simply Laplace transformation) is given by the integral

$$L[f(t)] = F(p) = \int_{-0}^{\infty} f(t)\, e^{-pt}\, dt \tag{3.199}$$

where p is the complex variable $p = a + if$,

f – frequency,

c – limit of convergence, i.e. the real component of the complex number $Re(p) = a \geqq c$.

The function $f(t)$ is assumed to equal zero for $t < 0$.

The symbol L [ ] is formally used for the Laplace transformation; the argument of this expression is called the *original* and by its transformation the *transform* is reached. The original $f(t)$ and its transform $F(p)$ form a unique transformation pair (Fig. 3.27).

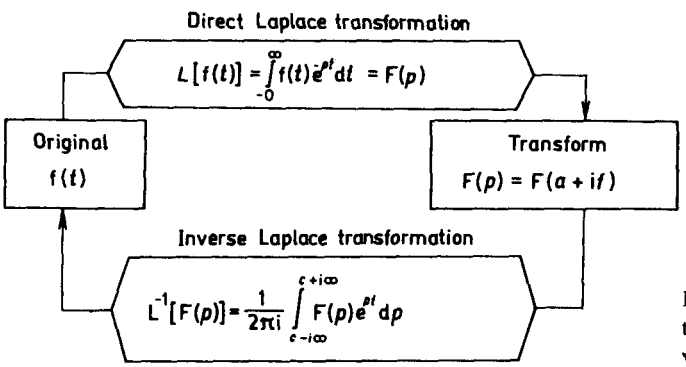


Fig. 3.27 Schematic representation of the direct and inverse Laplace transformation

The inverse Laplace transformation is an integral expression converting a function $F(p)$ back to the original time function $f(t)$ from which it came. The symbol L^{-1} [ ] is used for it

$$L^{-1}[F(p)] = f(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} F(p)\, e^{pt}\, dp \tag{3.200}$$

A function $f(t)$ is Laplace-transformable if:

(1) the function is piecewise continuous and unique in any time interval Δt (time functions are investigated),

(2) is has to be of exponential order, i.e. for each $\alpha > 0$ the following inequality has to be satisfied (for $M < \infty$)

$$|\lim_{t\to\infty} e^{-\alpha t} f(t)| \leqq M \tag{3.201}$$

(3) the real component of the complex number a in (3.199) has to be greater or equal to the limit of convergence c.

3.5.3 Laplace–Wagner Transformation

The direct Laplace–Wagner transformation is defined by the integral

$$L[f(t)] = p\int_0^{\infty} f(t)\, e^{-pt}\, dt = F(p) \tag{3.202}$$

and it gives p times greater transforms than the Laplace transformation. It has similar conditions of convergence to the Laplace transformation, viz. $Re(p) = a \geqq c$ where c is the limit of convergence.

Its application to technical problems is helpful as it retains the dimension of the original function. In the Laplace transformation the dimension of transformation is equal to the product of the dimension of the original function and time. The Laplace–Wagner transformation facilitates the checking of the dimension during the calculations.

The inverse Laplace–Wagner transformation is given by the expression

$$L^{-1}[F(p)] = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{F(p)}{p}\, e^{pt}\, dp = f(t) \tag{3.203}$$

The Laplace–Wagner transformation is formally identical with the calculus of operations which has been traditionally used in electrical engineering.

Example VIII

a) Calculate the Laplace transformation of the unit step function in the form

$$f(t) = \begin{cases} 1 & \text{if } t \geqq 0 \\ 0 & \text{if } t < 0 \end{cases} \tag{VIII.1}$$

Solution

Using (3.199) and assuming that $f(t) = 0$ for $t < 0$ we can write for $f(t) = 1$

$$\mathrm{L}[1] = F(p) = \int_0^\infty \mathrm{e}^{-pt}\,\mathrm{d}t = \int_0^\infty \mathrm{e}^{-(a+\mathrm{i}\omega)t}\,\mathrm{d}t = \frac{1}{a + \mathrm{i}\omega} = \frac{1}{p} \tag{VIII.2}$$

The value $1/p$ is the Laplace transform of the unit step function[1]).

b) Compute the Laplace–Wagner transformation for the same function

Solution

Using (3.202) we can write

$$\mathrm{L}[1] = F(p) = p\int_0^\infty \mathrm{e}^{-pt}\,\mathrm{d}t = p\,\frac{1}{p} = 1 \tag{VIII.3}$$

The Laplace–Wagner transformation of the unit step function is equal to one.

3.5.4 Basic Properties of the Laplace–Wagner Transformation

The basic properties of the Laplace–Wagner transformation for L-functions, i.e. Laplace transformable functions (see section 3.5.2) can be stated in a few points. Each of these properties is, in fact, a theorem, the proof of which can be found in specialized mathematical literature.

a) Transform of a sum of functions

For n functions $f_1(t)$, $f_2(t)$, ..., $f_n(t)$ and their Laplace–Wagner transforms $F_1(p)$, $F_2(p)$, ..., $F_n(p)$, the transform of the sum of these functions is equal to the sum of the transforms of each individual function, i.e. the Laplace–Wagner transformation is

$$\mathrm{L}[f_1(t) + f_2(t) + \ldots + f_n(t)] = F_1(p) + F_2(p) + \ldots + F_n(p) \tag{3.204}$$

This property is called the *theorem of linearity of transformation* and it is a consequence of its integral nature.

b) Transform of the function multiplied by a constant

For any constant the function $f_1(t) = \alpha f(t)$ possesses the transform $F_1(p) = \alpha\, F(p)$. Combining these two theorems, we get for the two functions $f(t)$ and $g(t)$ and the constants a, b

$$\mathrm{L}[a\, f(t) + b\, g(t)] = a\, F(p) + b\, G(p) \tag{3.205}$$

[1]) The unit step function does not possess e.g. Fourier transformation.

c) Transform of a function with change of time scale

For any $\beta > 0$,

$$L[f(\beta t)] = F\left(\frac{p}{\beta}\right) \tag{3.206}$$

This property is called the *similarity theorem of transformation.*

d) Transformation of delayed function

For each real $\delta > 0$ the function $f_1(t) = f(t - \delta)$ delayed a time δ is shifted to the right on the time axis and $f_1(t) = 0$ for $t \leqq \delta$. Its transform is

$$F_1(p) = L[f(t - \delta)] = e^{-p\delta} F(p) \tag{3.207}$$

For the function $f_2(t) = f(t + \delta)$ shifted to the left on the time axis the transform is

$$F_2(p) = L[f(t + \delta)] = e^{p\delta} F(p) \tag{3.208}$$

e) Transform of derivatives

Assume that $f(t)$ is continuous and differentiable in each interval Δt. Then the Laplace–Wagner transform of its first derivative with respect to time is

$$F^{(1)}(p) = L\left[\frac{d}{dt} f(t)\right] = p[F(p) - f(0)] \tag{3.209}$$

If all the derivatives of the function $f(t)$ at point $t = 0$ take on zero values then the transform of the n-th derivative is

$$F^{(n)}(p) = L\left[\frac{d^n}{dt^n} f(t)\right] = p^n F(p) \tag{3.209'}$$

f) Transform of integrals

Assume, as in the preceding case, that $f(t)$ is continuous and differentiable in each time interval Δt. Then the transform of its integral $\int_0^t f(u)\,du$ is

$${}^{(1)}F(p) = L\left[\int_0^t f(u)\,du\right] = \frac{1}{p} F(p) \tag{3.210}$$

Assuming that the n-dimensional integral exists for every n, that it is continuous and is of exponential order in the sense of the Laplace transformation convergence (section 3.5.2), then its transform is

$${}^{(n)}F(p) = \frac{1}{p^n} F(p) \tag{3.211}$$

g) Convolution of transforms

The transform of the product of two functions $f_1(t)$ and $f_2(t)$ with transforms $F_1(p)$ and $F_2(p)$, respectively, has the form

$$F(p) = L[f_1(t)\, f_2(t)] = \frac{p}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{1}{q} F_2(q) \frac{1}{p - q} F_1(p - q)\,dq \tag{3.212}$$

which can be written in a symbolic form

$$F(p) = F_1(p) * F_2(p) \tag{3.212'}$$

and is called *convolution of transforms.*

These properties can be effectively used for the computation of transforms of complicated functions without tedious calculation of the integrals that identify the transformations.

3.5.5 Use of Dictionary of Laplace Transforms for Inverse Transformation

In practical calculation of systems theory problems where the Laplace or Laplace–Wagner transformation is used, the most difficult part is the determination of the original function to the resulting transform defined by the inverse Laplace–Wagner transformation given by the general equation (3.203). Numerical calculation of this integral may be so difficult that the use of the Laplace–Wagner transformation would not be effective or advantageous. The simplest method of dealing with this problem is the application of the dictionary of Laplace–Wagner transforms, i.e. tables of the most frequent functions $f(t)$ and their transforms $F(p)$; they include also the basic properties and theorems of the Laplace–Wagner transformation. They can of course, be used for the determination of the inverse transformations.

The detailed dictionaries of integral transforms are extensive, and they are published separately (e.g. Erdélyi *et al.*, 1954) or are included in the specialized literature (e.g. Ditkin and Kuznyecov, 1954; Šalomoun, 1957). Table 3.16 provides an example of the headings in this dictionary and some of its functions and operations.

The first part includes the transform pairs for

- unit function,
- unit Dirac impulse,
- exponential function and its supplement to one,
- sine function.

In the second part the basic properties and theorems are listed for:

- the direct Laplace–Wagner transformation,
- the inverse Laplace–Wagner transformation,
- the Laplace–Wagner transformation of constants,
- linearity of Laplace–Wagner transformation,
- multiplication of the original function by a constant,
- similarity theorem for the Laplace–Wagner transformation.

Table 3.16 Sample of the dictionary of Laplace-Wagner transformations

I. Laplace–Wagner transformations	
original $f(t)$ $(t \geqq 0)$	transform $F(p)$
1	1
$\delta(t)$	p
$e^{\pm at}$	$p/(p \mp a)$
$1 - e^{\pm at}$	$a/(p \mp a)$
$\sin ft$	$fp\,(p^2 + f^2)$

II. Basic transform properties and theorems of the Laplace–Wagner transformations	
original $f(t)$ $(t \geqq 0)$	transform $F(p)$
$f(t)$	$p \int_0^\infty f(t)\, e^{-pt}\, dt$
$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{F(p)}{p} e^{pt}\, dt$	$F(p)$
c	c
$f_1(t) + f_2(t)$	$F_1(p) + F_2(p)$
$cf(t)$	$cF(p)$
$f(a/t)$	$F(p/a)$

3.5.6 Inverse Transformation by Heaviside Expansion

The Laplace–Wagner transformation often results in transforms in the form of a fraction of two polynomials of the complex argument p. The corresponding original function can be derived, in principle, by four methods; here the method of the residue theorem of the complex variable is used, which results in the Heaviside expansion.

Other methods e.g. partial fraction expansion with terms that are simple transforms, graphic solution of transforms and model technique seem to be more effetive only for special types of problems.

The principle of the inverse Laplace–Wagner transformation by the Heaviside expansion is the representation of the integral for the direct Laplace–Wagner transformation as the main value of the Cauchy's integral that is expressed by

$$f(t) = \frac{1}{2\pi i} \lim_{f \to \infty} \int_{c-if}^{c+if} \frac{F(p)}{p} e^{pt} \, dp \tag{3.213}$$

Assume that $F(p)$ is given by a fraction of two polynomials $M(p)$ and $N(p)$ in the form

$$F(p) = \frac{M(p)}{N(p)} = \frac{b_m p^m + b_{m-1} p^{m-1} + \ldots b_1 p + b_0}{p^n + a_{n-1} p^{n-1} + \ldots + a_1 p + a_0} \quad \text{for } (m \leqq n) \tag{3.214}$$

The expression $F(p)/p$ in e.g. (3.210) is then the fraction of two polynomials (3.214) with $m < n$ and the integral can be replaced by a summation in the form

$$f(t) = \sum_{i=0}^{m} \frac{M(p_i)}{p_i N'(p_i)} e^{p_i t} \tag{3.215}$$

which is called the *Heaviside expansion.* Parameters p_i called poles are the simple roots of the equation

$$p(p^n + a_{n-1} p^{n-1} + \ldots + a_1 p + a_0) = 0 \tag{3.216}$$

The original function is then determined for each term of expression (3.215) using the transformation linearity theorem and the dictionary of transform pairs. If the required pairs cannot be found in this dictionary, an expansion of the integrand to an infinite series is performed, which may result in simpler functions which can be found in it. If that is not the case, the integral representing the transformation is evaluated by graphic or numerical methods using computers.

Example IX

Determine the original function to the Laplace–Wagner transform in the form

$$F(p) = \frac{\omega p}{\omega^2 + p^2} \tag{IX.1}$$

Solution

If

$$F(p) = \frac{\omega p}{\omega^2 + p^2} = \frac{M(p)}{N(p)} \tag{IX.1'}$$

then equation (3.216) has the following form

$$p(p^2 + \omega^2) = 0 \tag{IX.2}$$

Its roots are

$$p_0 = 1$$
$$p_{1,2} = \sqrt{-1}\,\sqrt{\omega^2} \tag{IX.3}$$

hence

$$p_1 = \mathrm{i}\omega$$
$$p_2 = -\mathrm{i}\omega \tag{IX.3'}$$

Substituting them in equation (3.125) where $N'(p_i) = 2p_i$ we get the original function

$$\begin{aligned} f(t) &= \frac{M(p_1)}{p_1 \cdot 2p_1}\,\mathrm{e}^{p_1 t} + \frac{M(p_2)}{p_2 \cdot 2p_2}\,\mathrm{e}^{p_2 t} = \\ &= \frac{\omega p_1}{2p_1^2}\,\mathrm{e}^{p_1 t} + \frac{\omega p_2}{2p_2^2}\,\mathrm{e}^{p_2 t} = \frac{\omega}{2p_1}\,\mathrm{e}^{p_1 t} + \frac{\omega}{2p_2}\,\mathrm{e}^{p_2 t} = \\ &= \frac{\omega}{2\mathrm{i}\omega}\,\mathrm{e}^{\mathrm{i}\omega t} - \frac{\omega}{2\mathrm{i}\omega}\,\mathrm{e}^{\mathrm{i}\omega t} = \frac{1}{2\mathrm{i}}\,\mathrm{e}^{\mathrm{i}\omega t} - \frac{1}{2\mathrm{i}}\,\mathrm{e}^{-\mathrm{i}\omega t} = \\ &= \frac{\mathrm{e}^{\mathrm{i}\omega t} - \mathrm{e}^{-\mathrm{i}\omega t}}{2\mathrm{i}} = \sin \omega t \end{aligned} \tag{IX.4}$$

This can be simply checked. Using the form

$$\sin \omega(t) = \frac{\mathrm{e}^{\mathrm{i}\omega t} - \mathrm{e}^{-\mathrm{i}\omega t}}{2\mathrm{i}} \tag{IX.5}$$

and rearranging to

$$2\mathrm{i} \sin \omega(t) = \mathrm{e}^{\mathrm{i}\omega t} - \mathrm{e}^{-\mathrm{i}\omega t}$$

then

$$\begin{aligned} 2\mathrm{i}\,\mathrm{L}[\sin \omega(t)] &= \mathrm{L}[\mathrm{e}^{\mathrm{i}\omega t} - \mathrm{e}^{-\mathrm{i}\omega t}] = \mathrm{L}[\mathrm{e}^{\mathrm{i}\omega t}] - \mathrm{L}[\mathrm{e}^{-\mathrm{i}\omega t}] = \\ &= \frac{p}{p - \mathrm{i}\omega} - \frac{p}{p + \mathrm{i}\omega} = \frac{2\mathrm{i}\omega p}{p^2 + \omega^2} \end{aligned}$$

as $2\mathrm{i} \neq 0$ the expression

$$2\mathrm{i}\,\mathrm{L}[\sin \omega(t)] = \frac{2\mathrm{i}\omega p}{p^2 + \omega^2}$$

is

$$\mathrm{L}[\sin \omega(t)] = \frac{\omega p}{p^2 + \omega^2} \tag{IX.6}$$

and that is the transform of the desired function (IX.1).

3.5.7 Inverse Transformation by Convolution

The preceding examples illustrated the problem of the Laplace–Wagner transformation of a function, which can be expanded into the sum of simpler functions. In the control of complex systems there are functions with the transform in the form of a product of two or more transforms of simpler functions. In this situation the convolution principle can be applied.

Assume that functions $f_1(t)$ and $f_2(t)$ are Laplace-transformable L-functions and functions $F_1(p)$ and $F_2(p)$, respectively, are their transforms. Then it can be shown that

$$f(t) = \mathrm{L}^{-1}\left[\frac{F_1(p)\,F_2(p)}{p}\right] = \int_0^t f_1(t-u)\,f_2(u)\,\mathrm{d}u = f_1(t) * f_2(t) \tag{3.217}$$

where the notation $f_1(t) * f_2(t)$ means convolution of the original functions $f_1(t)$ and $f_2(t)$ and the integral

$$f(t) = \int_0^t f_1(t-u)\,f_2(u)\,\mathrm{d}u \tag{3.218}$$

is the convolution integral. The inverse transform of $F_1(p) \cdot F_2(p)$ is the function $g(t)$ that can be expressed as a derivation of the convolution integral with respect to t, i.e.

$$g(t) = \mathrm{L}^{-1}[F_1(p)\;F_2(p)] = \frac{\mathrm{d}}{\mathrm{d}t}\int_0^t f_1(t-u)\,f_2(u)\,\mathrm{d}u \tag{3.219}$$

The integral (3.218) is generally solved by the method per partes and there are four alternatives of its application to the function $g(t)$

$$\begin{aligned}
g(t) &= f_1(t)\,f_2(0) + \int_0^t f_1(u)\,\frac{\mathrm{d}f_2(t-u)}{\mathrm{d}(t-u)}\,\mathrm{d}u\,. \\
g(t) &= f_1(t)\,f_2(0) + \int_0^t f_1(t-u)\,\frac{\mathrm{d}f_2(u)}{\mathrm{d}u}\,\mathrm{d}u \\
g(t) &= f_1(0)\,f_2(t) + \int_0^t f_2(t-u)\,\frac{\mathrm{d}f_1(u)}{\mathrm{d}u}\,\mathrm{d}u \\
g(t) &= f_1(0)\,f_2(t) + \int_0^t f_2(u)\,\frac{\mathrm{d}f_1(t-u)}{\mathrm{d}(t-u)}\,\mathrm{d}u
\end{aligned} \tag{3.220}$$

The integrals (3.220) are called *Duhamel integrals*. The integral exists if the derivatives of $f_1(t)$ and $f_2(t)$ are continuous for $t > 0$. The values $f_1(0)$ and $f_2(0)$ are the values of functions $f_1(t)$ and $f_2(t)$ at time $t = 0+$.

3.6 Z-TRANSFORMATION

While the Fourier and Laplace transformations were developed for continuous time functions in system inputs and outputs, the z-transformation is a method intended for discrete sequences often used in hydrology and WRS problems (Delleur and Rao, 1971).

The z-transformation is defined by the transformation pairs:

$$X(z) = \sum_{n=0}^{\infty} x(n\,\Delta t)\, z^{-n} \tag{3.221}$$

and

$$x(n\,\Delta t) = \frac{1}{2\pi i} \oint X(z)\, z^{n-1}\, dz \tag{3.222}$$

where z is a complex variable, i is the imaginary unit, and $x(n\,\Delta t) = x(\Delta t)$, $x(2\,\Delta t)\ldots$ are the terms of the transformed sequence.

The line integral $\oint X(z)\, z^{n-1}\, dz$ is around the unit circle.

The z-transformation is sometimes referred to as an extension of the Laplace transformation, which is related to the possibility of its representation as a discretization of the Laplace integral by summation, which is formally identical with (3.221) with substitution of $z = e^{p\,\Delta t}$.

The transformation relationship between inputs and outputs is simpler for z-transformation than for Fourier or Laplace transformation. The z-transformation is considered in the form

$$Y(z) = H(z)\, X(z) \tag{3.223}$$

or

$$H(z) = \frac{Y(z)}{X(z)} \tag{3.224}$$

where $X(z)$ is the z-transformation of inputs $x(n\,\Delta t)$,
$Y(z)$ – the z-transformation of outputs $y(n\,\Delta t)$,
$H(z)$ – the transfer function.

3.7 BASIC NOTIONS OF CALCULUS OF VARIATIONS

3.7.1 Application of Calculus of Variations in WRS

Mathematical representation of reservoir release control and WRS problems can take the form of the *functional equation* or the *calculus of variations*. That is frequently the case with optimization of control based on Bellman's or Pontryagin's principle of optimality (Partl, 1968).

Variational equations in a general integral form for the optimization of reservoir or reservoir cascade operation used for hydroelectric power generation have been set up Kartvelishvili, 1970).

The calculus of variations is an important mathematical technique for technical and economic problems that otherwise could not be solved. It has penetrated into the world of WRS problems and has become useful tool for reservoir and WRS operation. As its implementation is desirable the principles of this method are outlined.

3.7.2 Functional

The calculus of variations is applied to the exact mathematical solution of optimization problems. One of its basic principles is the *functional* $I[y(x)]$. This can be interpreted as a generalization of the concept of the function where the independent variable is a certain function. Methods of finding the extreme of functions are well known. The calculus of variations investigates the possibilities and methods of functional minimum and maximum.

In general, the functional can be considered as a dependent variable with values determined by the choice of one or several functions from some particular class.

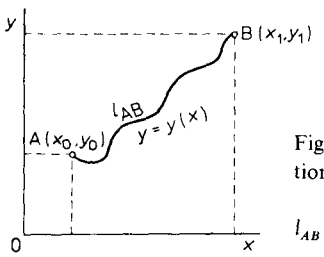


Fig. 3.28 Schematic representation of functional notion explanation

$$l_{AB} = I = I[y(x)] = \int_{x_0}^{x_1} \sqrt{1 + y'^2}\, dx$$

An example of the functional is one problem of determination of the segment l_{AB} of a curve in the plane connecting two terminal points A and B (Fig. 3.28). Assume that this curve is described mathematically by the function $y = y(x)$. The solution of this problem of the calculus of variations is the determination of such a function $y = y(x)$ so as to make l_{AB} minimum. This function is obviously a straight line

$$y = y_1 - \frac{y_1 - y_0}{x_1 - x_0}(x_1 - x) \tag{3.225}$$

3.7.3 Variations of Argument and Distance of Functions

Assume two functions $y(x)$ and $y_1(t)$ in a particular class of function, which can vary. Then the quantity

$$\delta y = y(x) - y_1(x) \tag{3.226}$$

which is formally some "difference function", is called the *increment* or *variation of argument* (variation of function) $y(x)$ of the functional $I[y(x)]$.

Functions $y(x)$ and $y_1(x)$ are at small distance in the sense of the distance of zero order (they are in the neighbourhood of zero order) if $|y(x) - y_1(x)|$ is small.

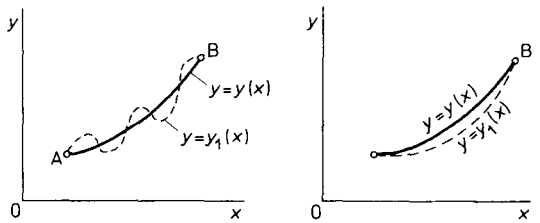


Fig. 3.29 Schematic representation of the distance of functions
a – zero order, *b* – higher orders

The distance of k-th order is defined as the maximum of: (1) difference $|y(x) - y_1(x)|$ in the interval $x \in \langle a, b \rangle$ where both functions are defined, and (2) the differences

$$|y'(x) - y_1'(x)|$$
$$\dots$$
$$|y^{(k)}(x) - y_1^{(k)}(x)|$$

where (k) is the k-th derivative[1]). A schematic representation of the distance of zero order and of higher orders is given in Fig. 3.29.

3.7.4 Some Properties of Functional

a) Continuity of functional

Continuity can be defined using the idea of distance of the k-th order. The functional $I[y(x)]$ is continuous on the curve $y_0(x)$ if $\delta > 0$ such that for each $\varepsilon > 0$ for all the functions $y(x)$ for which the functions are defined and that satisfy

$$\begin{aligned} &|y(x) - y_0(x)| < \delta \\ &|y'(x) - y_0'(x)| < \delta \\ &\dots \\ &|y^{(k)}(x) - y_0^{(k)}(x)| < \delta \end{aligned} \tag{3.227}$$

[1]) If the functions have small distances of the k-th order it is not necessary for the differences $|y^{(r)}(x) - y_1^{(r)}(x)|$ to be small for $r \geqq k + 1$.

the following holds:

$$|I[y(x)] - I[y_0(x)]| < \varepsilon \tag{3.228}$$

b) Linearity (homogeneity) of functional

A functional is linear (homogeneous) if it satisfies the conditions:

$$I[cy(x)] = cI[y(x)] \tag{3.229}$$

(c is an arbitrary constant) and

$$I[y_1(x) + y_2(x)] = I[y_1(x)] + I[y_2(x)] \tag{3.230}$$

c) Variation of functional

The concept of variation of a functional is based on the idea of variation of an argument (function). The increment of the argument is

$$\delta y = y(x) - y_1(x) \tag{3.231}$$

and the corresponding increment of the functional is

$$\delta I = I[y(x) + \delta y] - I[y(x)] \tag{3.232}$$

which can be written in the form

$$\delta I = I[y(x); \delta y] + \beta[y(x); \delta y] \max |\delta y| \tag{3.233}$$

The first term $I[y(x); \delta y]$ is a linear (homogeneous) functional in δy and is called the *variation of the functional* $I[y(x)]$ or the main component of the increment of the functional.

The second part of the expression for the increment of the functional given, the functions $\beta[y(x); \delta y]$ converges to zero if the distance $\max |\delta y| \to 0$. It is obvious that the variation of the functional is analogous to the differential of functions.

Define function $\varphi(\alpha)$ of one variable (parameter) α:

$$\varphi(\alpha) = I[y(x) + \alpha \delta y] \tag{3.234}$$

where $y(z)$ and δy are constants.

The variation of the functional $I[y(x)]$ can be expressed as the derivation of the function (3.234) with respect to α at the point $\alpha = 0$, i.e.

$$\frac{\partial}{\partial \alpha} I[y(x) + \alpha \delta y]_{\alpha = 0} \tag{3.234'}$$

d) Extrema of functional

The functional assumes an extremal value for a function $y = y_0(x)$ if the corresponding variation of functional δI is equal to zero. The notion of the extrema of the functional can be specified and made precise by the order of the distance of curves $y(x)$ and $y_0(x)$.

The functional $I[y(x)]$ takes on its *maximum value* if its values are not higher on any curve $y = y(x)$ in the neighbourhood of the curve $y = y_0(x)$ than on this curve, i.e.

$$\delta I = I[y(x)] - I[y_0(x)] \leqq 0 \tag{3.235}$$

If the increment of functional δI is equal to zero for the functions $y = y_0(x)$ only, then this maximum is called a strict maximum.

On the other hand, the functional $I[y(x)]$ takes on its *minimum value* for the function $y_0(x)$, if $\delta I \geqq 0$. The extrema of functionals can be classified further into weak and strong.

The functional possesses a *weak extremum* if it occurs on a curve $y = y_0(x)$, which has a small distance of the first order for all curves $y = y(x)$.

The functional possesses a *strong extremum* if it occurs on a curve $y = y_0(x)$ which has a small distance of zero order for all the curves $y = y(x)$. It can be concluded that a necessary condition for the existence of the extremum of a functional $I[y(x)]$ of the function $y = y_0(x)$ is satisfaction of the relationship

$$\frac{\partial}{\partial \alpha} I[y_0(x) + \alpha \delta y] \big|_{\alpha=0} = 0 \tag{3.236}$$

Hence

$$\frac{\partial}{\partial \alpha} I[y(x, \alpha)] \big|_{\alpha=0} = 0 \tag{3.237}$$

where $y(x, \alpha)$ is a family of curves starting with the curve $y = y_0(x)$ for $\alpha = 0$ and ending with the curve $y = y_0(x) + \delta y$ for $\alpha = 1$.

e) Functionals of several functions and multivariate functions

These types of functional have the form

$$I = I[y_1(x), y_2(x), \ldots, y_n(x)] \tag{3.238}$$

and

$$I = I[z_1(x_1, x_2, \ldots, x_n); z_2(x_1, x_2, \ldots, x_n); \ldots; z_m(x_1, x_2, \ldots, x_n)] \tag{3.239}$$

For example, for the functional $I = I[z(x, y)]$ the extremum can be found using the condition

$$\frac{\partial}{\partial \alpha} I[z(x, y) + \delta z] = 0 \tag{3.240}$$

where $\delta(z)$ is the increment of the function $z = z(x, y)$.

3.7.5 Two-Fixed-Point Boundary-Value Variation Problem

The simplest type of functional is

$$I[y(x)] = \int_{x_0}^{x_1} F\left[x, y(x), \frac{dy(x)}{dx}\right] dx \tag{3.241}$$

Assume that the feasible functions $y(x)$ that satisfy this functional have their two terminal points fixed on the abscissa x_0 and x_1, and that the function F is differentiable up to the third order. The problem of finding the extremum of this functional is called the two-fixed-point boundary-value problem in contrast to some problems where there is some degree of freedom for boundary values, and which are more complicated and more difficult to solve.

The extremum of functional (3.241) can be found on a curve $y = y(x)$ where $\delta \boldsymbol{I} = 0$.

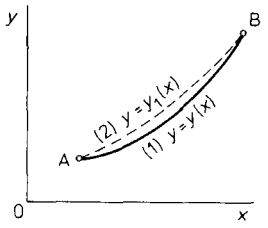


Fig. 3.30 The relative position of the desired (*1*) and the compared (*2*) curves in computation of the extremal curve for the two-point boundary-value problem

Assume some feasible curve in small distance of $y = y(x)$ in the sense of the distance of the k-th order $k > 0$ (see Fig. 3.31). Both these curves $y(x)$ and $y_1(x)$ belong to a parametrized family of curves $y(x, \alpha)$ with one parameter α, i.e.

$$y(x, \alpha) = y(x) + \alpha[y_1(x) - y(x)] \tag{3.242}$$

where $y_1(x) - y(x) = \delta y$ is the variation of function $y(x)$.

In this one-parameter system of curves the desired curve $y(x)$ can be found for $\alpha = 0$ where the functional $\boldsymbol{I}[y(x)]$ assumes its maximum, and the corresponding curve $y_1(x)$ for $\alpha = 1$ (Fig. 3.30).

In the following calculation, instead of the functional $\boldsymbol{I}[y(x)]$, the function $\varphi(\alpha)$ of the parameter α, defined by equation (3.234) is investigated. The necessary condition for the existence of the extremum at point $\alpha = 0$ is $\mathrm{d}\varphi/\mathrm{d}\alpha|_{\alpha=0} = 0$, which can be written for function $\varphi(\alpha)$ as

$$\varphi(\alpha) = \int_{x_0}^{x_1} F\left[x, y(x, \alpha), \frac{\partial y}{\partial x}(x, \alpha)\right] \mathrm{d}x \tag{3.243}$$

Differentiating $\mathrm{d}\varphi/\mathrm{d}\alpha$ and letting $\mathrm{d}\varphi/\mathrm{d}\alpha|_{\alpha=0} = 0$ we get

$$\int_{x_0}^{x_1} \left[F_y \delta_y + F'_y \delta \left(\frac{\partial y}{\partial x}\right)\right] \mathrm{d}x = 0 \tag{3.244}$$

where

$$F_y = \frac{\partial}{\partial y} F\left[x, y(x, \alpha), \frac{\partial y}{\partial x}(x, \alpha)\right]$$

and

$$F'_y = \frac{\partial}{\partial\left(\frac{\partial y}{\partial x}\right)} F\left[x, y(x, \alpha), \frac{\partial y}{\partial x}(x, \alpha)\right] \tag{3.245}$$

Equation (3.244) can be further simplified to

$$\int_{x_0}^{x_1}\left[F_y - \frac{\mathrm{d}}{\mathrm{d}x} F'_y\right] \delta y \, \mathrm{d}x = 0 \tag{3.246}$$

Using the principal theorem of the calculus of variations[1]) in equation (3.246),

$$F_y - \frac{\mathrm{d}}{\mathrm{d}x} F'_y = 0 \tag{3.248}$$

Equation (3.248) is the *Euler equation*, one of the principal equations which have to be solved in the calculus of variations. It can be written in a general form

$$y = y(x, C_1, C_2) \tag{3.249}$$

where the constants C_1 and C_2 are determined from the initial conditions

$$y(x_0) = y_0; \quad y(x_1) = y_1 \tag{3.250}$$

The integral curves (3.249) give the solution of the problem as they are extremal curves of the functional of the type (3.241).

Example X

Given the functional

$$I[y(x)] = \int_0^{\pi/2} [(y')^2 - y^2] \, \mathrm{d}x \tag{X.1}$$

with the boundary values

$$y(0) = 0; \quad y\left(\frac{\pi}{2}\right) = 1 \tag{X.2}$$

Find the function where it assumes its extremum.

---

[1]) If the function $\Phi(x)$ is continuous in the interval $\langle x_0; x_1 \rangle$ and if

$$\int_{x_2}^{x_1} \Phi(x) \, \eta(x) \, \mathrm{d}x = 0 \tag{3.247}$$

then for any function $\eta(x)$ satisfying any general conditions (e.g. differentiability to the k-th order and given boundary the values x of the interval etc.), for $x \in \langle x_0; x_1 \rangle$,

$$\Phi(x) = 0 \tag{3.247'}$$

Solution

Functional (X.1) is of the type (3.241). Therefore the following partial derivatives are calculated

$$F_y = \frac{\partial}{\partial y} F(x, y, y') = -2y$$

$$F'_y = \frac{\partial}{\partial y'} F(x, y, y') = 2y' \tag{X.3}$$

and the Euler equation is written in the form

$$F_y - \frac{\mathrm{d}}{\mathrm{d}x} F_{y'} = 0 \tag{X.4}$$

$$-2y - \frac{\mathrm{d}}{\mathrm{d}x}(2y') = 0$$

hence

$$y'' + y = 0 \tag{X.5}$$

The solution of the differential equation (X.5) is the function

$$y = C_1 \cos x + C_2 \sin x \tag{X.6}$$

Using the boundary values (X.2) we get

$$C_1 = 0; \quad C_2 = 1 \tag{X.7}$$

and the extremal function of the function to be found is

$$y = \sin x \tag{X.8}$$

3.7.6 Principle of Two-Free-Point Boundary-Value Variation Problem

In these problems it is assumed that one or both terminal points $[x_0, y_0]$; $[x_1, y_1]$ in a two-dimensional problem are free. The class of feasible functions including the extremal function is thus enlarged. The function to be found is again the solution of the Euler equation, which contains two free constants since the differential equation is of the second order. In a two-free-point problem one or both boundary-value conditions (3.250) are not applicable. The values of the constants are calculated using condition that variation of the functional has to be equal to zero, i.e. $\delta I = 0$.

Assume a fixed point $[x_0, y_0]$. Then all extremal functions will pass through this point and form a bunch of curves (Fig. 3.31a).

Assume that the point $[x_1, y_1]$ moves on the curve

$$y_1 = \varphi(x_1) \quad \text{(see Fig. 3.31b)} \tag{3.251}$$

In such a case, an additional condition is applied in the form of approximate equality of the variation of the boundary values δy_1 and the differential of the function $\varphi(x_1)$, i.e.

$$\delta y_1 \approx \varphi'(x_1)\,\delta(x_1) \tag{3.252}$$

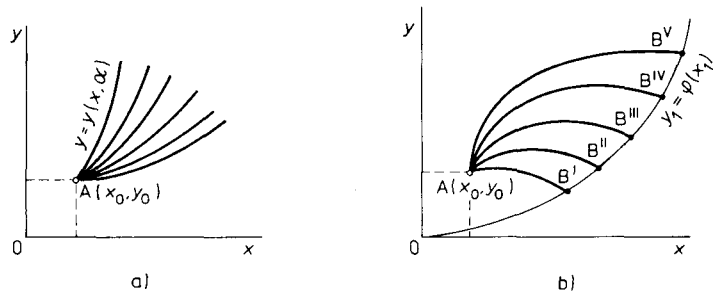


Fig. 3.31 Bunch of extremal curves in the two-free-points boundary-value problem
a – fixed terminal point $[x_0, y_0]$, *b* – fixed terminal point $[x_0, y_0]$ and the point $[x_1, y_1]$ on a curve $y_1 = \varphi(x_1)$

3.7.7 Application of Calculus of Variations for Stochastic Optimization Problems

Use of the calculus of variations for optimization problems and the differences in its use for deterministic and stochastic optimization problems are illustrated by a case study of water resources and hydroelectric systems.

Kartvelishvili, 1970, described a case study of the long-term operating policy of a single-purpose reservoir used for hydroelectric power generation in a system of thermal and water power plants.

The continuity equation relating reservoir inflow $Q_1(t)$, release $Q_2(t)$, quantity of water in reservoir W and any change in that quantity in the form of a derivative $\mathrm{d}W/\mathrm{d}t$ can be written in the form

$$Q_2(t) = Q_1(t) - \frac{\mathrm{d}W}{\mathrm{d}t} \tag{3.253}$$

If no spills occur (all the release passes through the turbines) the average power production of the water power plant in a particular time interval is the function

$$N(t) = N[Q_2(t); W] = N\left[Q_1(t) - \frac{\mathrm{d}W}{\mathrm{d}t};\ W\right] \tag{3.254}$$

where the change of the reservoir storage W also determines the change of the head of the water power plant. If the total load of the power system in time t is $S(t)$ then the load of the thermal power plants is

$$N_T(t) = S(t) - N\left[Q_1(t) - \frac{dW}{dt};\ W\right] \tag{3.255}$$

Normal operation of the power system is assumed where deficits can occur due to low flows (water power plants) or due to insufficient supplies of fuel (thermal plants).

In any time interval $\langle t_1, t_2\rangle$ a proportion of the losses is due to reservoir operation policy as given by the reliability theory (Ireson, 1969, Mirtskhulava, 1974) and it can be expressed by the relationship

$$Z_1 = \int_{t_1}^{t_2} e^{-p(t-t_1)} B\, dt \tag{3.256}$$

where $p = 1/T$,

T – is the economic lifetime,

B – loading function of the thermal power plant in the power system

$$B = B\left\{S(t) - N\left[Q_2(t) - \frac{dW}{dt};\ W\right]\right\} \tag{3.257}$$

or, in brief,

$$B = B\left[t, Q_2(t), W, \frac{dW}{dt}\right] \tag{3.257'}$$

The expectation of losses is minimized. The solution of the optimization problem taken as a deterministic one is the best operation policy $W = W(t)$, sometimes called the operation strategy. The expression (3.255) is the functional equation.

If the stochastic approach is used for optimization, then the formulation of the problem must be modified. The extremal function that has been found for time t_1 determines the operation policy for the whole time interval $\langle t_1, t_2\rangle$. In the deterministic model the same information is used throughout the whole interval $\langle t_1, t_2\rangle$. In the stochastic model the stage of the system can change at each time $t \in \langle t_1, t_2\rangle$. At the point t_1 a prediction of the inflow $Q_1(t)$ is possible. By means of this and hydrometeorological, water management and balance data, the optimal operation policy can be selected; it is denoted by $W_{t,1}(t)$. At a chosen time $t' > t_1$ an improvement of the operation policy $W_{t,1}(t)$ to $W_{t'}(t)$ may be carried out and its use assumed in the remainder of the time interval $(t_2 - t')$. The new operation policy $W_{t'}(t)$ should better satisfy the conditions of minimum losses given by the functional equation that can be derived for this case in the form

$$Z_1' = \int_{t_1}^{t_2} e^{-p(t-t_1)} \left\{ M\left[\frac{\partial B}{\partial W}\right] + pM\left[\frac{\partial B}{\partial\left(\frac{dW}{dt}\right)}\right] - M\left[\frac{d}{dt}\,\frac{\partial B}{\partial\left(\frac{dW}{dt}\right)}\right] \right\} dt\, \delta M \tag{3.258}$$

where δW is the variation of the function $W(t)$, M [ ] is the expectation (mean value) of the variable analysed.

This approach also leads to the functional equation, i.e. to the solution of a variation problem. This method, however, has not yet been implemented practically in WRS problems.