proportion of observations from this distribution would exceed 7.0? Plot the probability density function.

5.29 From a data series of minimum annual discharges on a stream, one obtained an average of 175 cfs, a standard deviation of 65 cfs, and a coefficient of skew of 1.5. Using both extreme value type I (for) minimum and type III (for minimum) distributions, evaluate the probability of an annual minimum flow being less than 125 cfs.

5.30 If flood flows from a large watershed have an average value of 1,500 cms with a variance of 53,500 $(\text{cms})^2$, what is the probability that a flood will be equal to or exceed 2,000 cms, if the Gumbel distribution is used?

Chapter 6

Impulse Response Functions as Probability Distributions

A multitude of environmental and hydrologic processes embody both the elements of chance and the descriptive laws of physics. A finer process description at one scale is lost through the processes of integration in time and space and through averaging. This justifies simplification in representation of the processes. It is hypothesized that if an environmental process is described by a linear or linearized governing equation, then the solution of this equation for a unit impulse (or Dirac delta) function can be interpreted as a probability density function for describing the probabilistic properties of the process. This hypothesis is tantamount to mapping from the unit impulse response (UIR) function, $h(t)$, to the probability density function (PDF), $f(x)$, where h is the UIR as a function of time or space variable denoted by t and f is the PDF as a function of the random variable of the process. For example, the impulse response of a diffusion equation for pollutant transport described by the space–time variation of concentration can be used as a probability distribution for pollutant concentration in a medium, such as a river, a lake, conduit storm water, soil, or a saturated geologic formation. Likewise, the impulse response of a linearized diffusion model of channel flow can be interpreted as a probability distribution for frequency analysis of extreme values (such as floods, droughts, hurricanes, earthquakes, and so on). Similarly, the impulse response of a linear reservoir can be used as an exponential probability distribution model. The impulse response of a cascade of equal linear reservoirs is the gamma distribution, which has a number of applications in environmental and water resources data analysis. In this vein, a number of impulse responses of physically based equations that apply to environmental and hydrologic processes and data are discussed and illustrated by using field or laboratory data.

Environment can be defined as a continuum of three components: air, water, and land. The processes dealing with these components and their dynamic interactions constitute environmental processes. Examples of such processes are solute transport in a river, a lake, storm water, soil, or an aquifer; flood; drought; rainfall; erosion; sediment transport by storm water or in a river; air pollution; depletion in the ozone layer; glacial movement and melting; climate change; occurrence of an epidemic resulting from pollution; seawater rise; and salt-water intrusion. A quantitative description of these processes involves a determination of the space, time, or space–time history of flux, concentration, the peak, the average, or the volume of the process variable. For example, to describe the transport of a pollutant in a river, the pollutant flux or concentration as a function of space and time may be selected. One may also select the pollutant load passing through a given point on the river over a selected period of time, say, a month or a year. Similarly, to describe the quality of air in an urban area, one may select the ozone level and determine its variability in time.

Because environmental processes frequently embody both the elements of chance and the descriptive laws of physics, environmental variables cannot be completely described either by deterministic means or by stochastic means alone. Rather, a better approach has to be a combination of both the stochastic and the deterministic means. Considerable simplification is usually needed to view the variables deterministically. This can be justified in light of the observation that excessive process description at one scale is lost through the operations of integration in time and averaging. Furthermore, the governing equations themselves have inherent limitations with regard to accuracy. Even more stark is the state of data acquisition and processing.

When the stochastic aspect of the environmental variables is considered, the element of chance is attributed to a complex mix of factors, such as inherent stochasticity in environmental processes owing to interactions of the environmental continuum components, human–environment interaction, and our limitations to observe and quantify spatial and temporal variability of environmental variables. A stochastic description usually includes analysis of variance, time series analysis, or frequency analysis. The type of analysis that is needed depends on the demand of the problem. For example, a frequency analysis is needed for planning and design. A time series analysis is needed for operation and management. A regression analysis is needed for prediction, extrapolation, or interpolation. Analogous to a deterministic description, a stochastic analysis also involves simplifications that can be justified on the basis of lack of data or lack of adequate knowledge of processes to be modeled as well as of the methodologies for modeling of nonlinear and non-Gaussian processes, requirement of simplicity, parsimony of parameters, and so on.

An environmental process in nature exhibits itself in many ways and a variable selected to describe some aspect of this process must obey the commands of the process. The variable has an intrinsic nature and its characterization and analyses, whether deterministic or stochastic or a combination thereof, is important. This means that there must be an inherent connectivity among these analyses. In environmental science and engineering, this connection is frequently observed. For example, a regression analysis without error analysis in statistics is no different from curve-fitting techniques in mathematics. Indeed, regression analysis techniques are often employed to find a best-fit curve for a given set of data, and the connection between these types of techniques is well known. Another example is the autoregressive (AR) technique in time series analysis. When an AR model is applied to, say, daily, monthly, or annual river flow, the coefficients associated with the autoregressive variables are nothing but the ordinates of linear kernel of the flow variable. Since the AR technique is linear, it is equivalent to the impulse response function of a linear flow process. In hydrology, this is known as the unit hydrograph method. One also finds a connection between the unit hydrograph method and spectral analysis. This means that certain linear time series analysis techniques are equivalent to linear response functions of environmental processes. However, the connection between frequency analysis methods and deterministic methods is not clear yet. This may be because frequency, by definition, is stochastic in nature and finding a deterministic equivalent seems somewhat contradictory in terms. However, our objective here is to find a connection through techniques of analysis, not through concept. This constitutes the subject matter of this chapter.

6.1 Hypothesis

It is hypothesized that if an environmental variable is described by a linear or linearized governing equation, then the solution of this equation for a unit impulse (or Dirac delta) function (UIF) can be interpreted as a PDF for describing the probabilistic properties of the random variable, say, X. The solution for the UIF can be characterized as the UIR or $h(t)$. If the UIR is a function of time t, then the PDF is a mapping from the (h, t) plane to the (f, x) plane, where x is the value (or quantile) of the random variable X for which $h(x, t)$ is desired, and f is the PDF.

There are many environmental variables that can be reasonably well described linearly. If some of the variables cannot be described linearly in the real domain, then they can be described linearly in the logarithmic domain or in an appropriate transformed domain. Examples of linear approximation are surface runoff from rainfall excess, river flow, monthly sediment discharge, and solute concentration in a tube or soil. Thus, their UIRs can be considered as their PDFs. It is not surprising that several probability distributions have found their niche in linear environmental analyses. This hypothesis will be explored in what follows.

6.2 Impulse Responses of Linear Systems

6.2.1 Linear Reservoir

The simplest linear system in hydrology is probably a linear reservoir (or storage element), shown in Fig. 6-1, and is described by the spatially lumped form of the continuity equation

$$\frac{dS(t)}{dt} = I(t) - Q(t) \tag{6.1}$$

and a storage–discharge type relation

$$S = k\,Q \tag{6.2}$$

where $I(t)$ is the rate of inflow to the reservoir at time t, $Q(t)$ is the rate of outflow from the reservoir at time t, $S(t)$ is the storage in the reservoir at time t, and k is the storage coefficient or average travel (or residence or lag) time. Substitution of Eq. 6.2 in Eq. 6.1 yields

$$I(t) - Q(t) = \frac{dS(t)}{dt} = k\frac{dQ}{dt} \tag{6.3}$$

Solution of Eq. 6.3 gives

$$Q_t = I[1 - \exp(-t/k)] \qquad t \le D \tag{6.4}$$

and

$$Q_t = Q_P \exp[-(t-D)/k] \qquad t \ge D \tag{6.5}$$

where D is the inflow duration, and Q_P is the peak of outflow hydrograph given by

$$Q_P = I[1 - \exp(-D/k)] \tag{6.6}$$

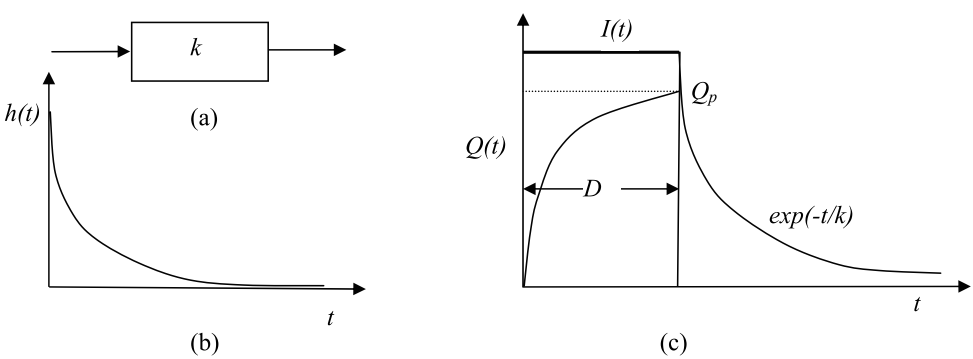

Figure 6-1 *Depiction of linear reservoir concept: (a) lag time, (b) IUH, and (c) hydrograph due to a pulse of* D *hour duration.*

The linear reservoir has been used for rainfall–runoff modeling either by itself or as an element of a network model. For instantaneous inflow that fills the reservoir in time $t = 0$,

$$Q = \frac{S}{k} \exp(-t/k) \tag{6.7}$$

If $I(t)$ is denoted by a unit delta function $\delta(t)$, then the UIR of the linear reservoir, $h(t)$, is

$$h(t) = \frac{\exp(-t/k)}{k} \tag{6.8}$$

In hydrology, $h(t)$ is known as the instantaneous unit hydrograph (IUH). The determination of $h(t)$, the impulse response function (or the kernel function or Green's function) of a system from input and output data, is known as system identification. Convolution of the impulse response function with the system inputs gives the system output. Then, the PDF of a variable described by a linear reservoir becomes

$$f(x) = \frac{\exp(-x/k)}{k} \tag{6.9}$$

where x is the quantile of the variable X described by the linear reservoir and k is a parameter. Thus, it is seen that $h(t)$ is mapped onto $f(x)$. Equation 6.9 is an exponential density function and is widely used in environmental and water resources. For example, if an environmental process is described by the Poisson process then the interarrival times follow an exponential distribution. Interarrival times of floods can be modeled by using Eq. 6.9. Rainfall depth, intensity, and duration have been modeled with Eq. 6.9. It should be noted that k in $f(x)$ represents the average of X and hence its interpretation from Eq. 6.8 remains unchanged under mapping of $h(t)$ onto $f(x)$.

Another modification of the linear reservoir involves restating the unit delta function $\delta(t)$ as $\delta(t - t_0)$, where t_0 is the time at which the function occurs. In that case, $h(t)$ of Eq. 6.8 becomes

$$h(t) = \frac{\exp[-(t - t_0)/k)}{k} \tag{6.10}$$

Equation 6.10 is the UIR of a lag and route linear reservoir system in which t_0 is the amount of lag time before water is released from the reservoir. This is equivalent to a linear reservoir and a linear channel, connected in series. By mapping Eq. 6.7 onto the probability plane, the PDF becomes

$$f(x) = \frac{\exp[-(x - x_0)/k)}{k} \tag{6.11}$$

where x_0 is the threshold of X, $x \geq x_0$. The threshold is the minimum value of X. This is useful in frequency analysis of environmental data.

Example 6.1 Consider a linear reservoir with a lag time of 10 hours. This reservoir receives a pulse of 10 m^3/s for a duration of 5 hours. Determine the peak outflow and graph the outflow hydrograph. Also graph the outflow hydrograph when lag time is 5 hours and compare the two hydrographs.

Solution The peak of the hydrograph can be computed from Eq. 6.6:

$$Q_P = I[1 - \exp(-D/k)] = 10[1 - \exp(-5/10] = 3.93 \text{ cumec}$$

The hydrograph for both cases is plotted in Fig. 6-2. Notice that the peak is higher when the lag time is smaller and the recession is slower and lasts longer when k is larger. Clearly, a larger catchment will have a longer lag time. This hydrograph can also be considered as a probability density function of peak discharge exceeding a given threshold.

6.2.2 Muskingum Model

The Muskingum model is described by Eq. 6.1 and the Muskingum hypothesis:

$$S(t) = K\left[\alpha I(t) - (1-\alpha)Q(t)\right] \tag{6.12}$$

where K is the average reach travel time and α is a parameter or a weighting coefficient. The unit impulse response of the Muskingum model (see Fig. 6-3), is given by

$$h(t) = -\frac{\alpha}{1-\alpha}\delta(t) + \frac{1}{K(1-\alpha)^2}\exp\left[-\frac{t}{K(1-\alpha)}\right] \tag{6.13}$$

It has been shown that modeling flood routing along a short reach of a lowland river may result in the negative value of α and

$$0 \le \left(-\frac{\alpha}{1-\alpha}\right) \le 1 \tag{6.14}$$

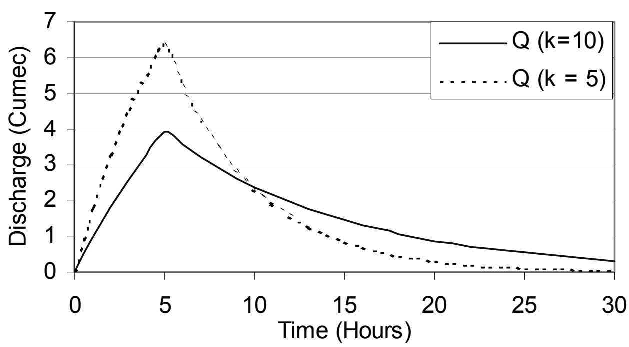

Figure 6-2 Outflow hydrograph from the linear reservoir for two values of lag time.

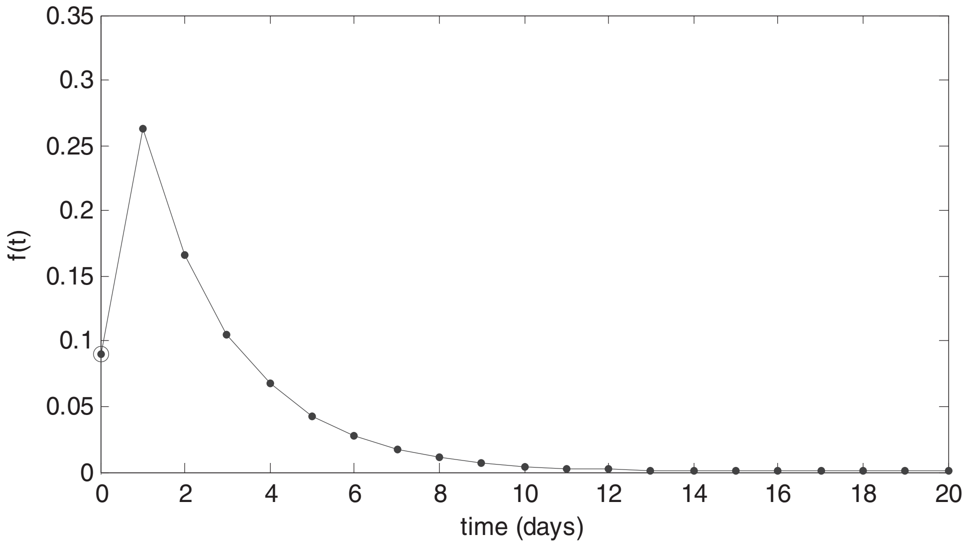

Figure 6-3 *Muskingum reach impulse response function.*

Denoting $\frac{1}{1-\alpha} = \beta$ and $\frac{\beta}{K} = \gamma$ and renaming t as x, one gets a two-parameter probability distribution function:

$$f(x) = (1 - \beta)\delta(x) + \beta\gamma \exp(-\gamma x) \tag{6.15}$$

The PDF given by Eq. 6.15 is a weighted sum of two functions: a delta function and an exponential function. It is interesting to note that in this function parameter β is a weighting factor and parameter $K = \beta/\gamma$ becomes the average of X. Thus, the original expressions of the weighting factor and the average travel time are modified under mapping, but the conceptual meaning of the modified expressions remains more or less intact. Equation 6.15 is useful for frequency analysis of floods with zero values as well as flood damage.

Example 6.2 Let the average travel time of a Muskingum lowland reach, K, be 2 days, and the weighting coefficient parameter be -0.1. Determine the impulse response function of reach outflow.

Solution According to Eq. 6.15,

$$\beta = \frac{1}{1+0.1} = \frac{1}{1+0.1} = 0.91, \gamma = \beta / K = 0.91/2 = 0.455$$

Then

$$f(t) = 0.09\delta(t) + 0.414 \exp(-0.455t)$$

6.2.3 Cascade of Linear Reservoirs

If an environmental system is represented by a cascade of n equal linear reservoirs, then its UIR becomes

$$h(t)=\frac{1}{k\Gamma(n)}\left(\frac{t}{k}\right)^{n-1}\exp(-t/k) \quad \textbf{(6.16)}$$

where k is the storage parameter of each reservoir and $\Gamma(n)$ is the gamma of n. Since there are n reservoirs, nk represents the total lag time (or the average residence time) of the system. By mapping onto the probability plane, the PDF becomes

$$f(x)=\frac{1}{k\Gamma(n)}\left(\frac{x}{k}\right)^{n-1}\exp(-x/k) \quad \textbf{(6.17)}$$

where k and n are parameters. Equation 6.17 is the gamma probability density function. The gamma distribution results from the sum of exponentials, where n is the number of exponentials. In deterministic parlance, each exponential represents a linear reservoir. Thus, the deterministic interpretation of parameters is carried over through mapping. The gamma distribution is one of the most commonly used probability distributions for environmental frequency analysis.

If an environmental system satisfies the requirement that $h(t) > 0$ if $t \geq t_0$ then the UIR becomes

$$h(t)=\frac{1}{k\Gamma(n)}\left(\frac{t-t_0}{k}\right)^{n-1}\exp[-(t-t_0)/k)] \quad \textbf{(6.18)}$$

The interpretation of t_0 is that the cascade of equal linear reservoirs retains water for time t_0 before it starts to release it. Mapping onto the probability plane transforms Eq. 6.18 into

$$f(x)=\frac{1}{k\Gamma(n)}\left(\frac{x-x_0}{k}\right)^{n-1}\exp[-(x-x_0)/k)] \quad \textbf{(6.19)}$$

which represents the three-parameter Pearson type III probability density function. This is equivalent to a cascade of linear reservoirs and channels connected in series. This is one of the most widely used frequency distributions in hydrology and environmental sciences. Note that Eq. 6.17 is a special case of Eq. 6.19. Here parameter x_0 is the lowest value or threshold of the variable X. Although these parameters, k, n, and x_0, can be interpreted by using the deterministic analogy, their optimal values are better found by curve fitting. This means that under mapping onto the probability plane, the interpretation of the parameters may be somewhat transformed.

Example 6.3 Take n as 3 and k as 6 hours. Compute the probability density function of peak discharge. Assume the lowest value or threshold of discharge as 100 cumecs.

Solution Substituting $n = 3$, $k = 6$, and $x_0 = 100$ into Eq. 6.19, we have the probability density function of discharge as

$$f(x)=\frac{1}{6\Gamma(3)}\left(\frac{x-100}{6}\right)^2 \exp[-(x-100)/6)]$$

The probability density function is plotted in Fig. 6-4.

6.2.4 Linear Downstream Channel Routing Model

One of the most important problems in one-dimensional flood routing analysis is the downstream problem (i.e., the prediction of flood characteristics at a downstream section on the basis of the knowledge of flow characteristics at an upstream section). By using the linearization of the Saint-Venant equation, the solution of the upstream boundary problem was derived by Deymie (1939), Masse (1937), Dooge and Harley (1967), and Dooge et al. (1987a,b), among others; a discussion of this problem is presented in Singh (1996a). The solution is a linear, physically based model with four parameters dependent on the hydraulic characteristics of the channel reach at the reference level of linearization. However, the complete linear solution is complex in form and is relatively difficult to compute (Singh 1996a). Two simpler forms of the linear channel downstream response are recognized in the hydrologic literature and are designated as the linear diffusion (LD) analogy model and the linear rapid flow (LRF) model. These correspond to the limiting flow conditions of the linear channel response, that is, where the Froude number is equal to zero (Hayami 1951; Dooge 1973) and where it is equal to one (Strupczewski and Napiorkowski 1980c).

The complete linearized Saint-Venant equation is of hyperbolic type and may be written as

$$a\frac{\partial^2 Q}{\partial y^2}+b\frac{\partial^2 Q}{\partial y \partial t}+c\frac{\partial^2 Q}{\partial t^2}=d\frac{\partial Q}{\partial y}+e\frac{\partial Q}{\partial t} \tag{6.20}$$

where Q is the perturbation of flow about an initial condition of steady-state uniform flow, y is the distance from the upstream boundary, t is the elapsed time, and a, b, c, d, and e are parameters as functions of channel and flow characteristics at the reference steady-state condition. A number of models of simplified forms of the complete Saint-Venant equation have been proposed in the hydrological literature.

If all three second-order terms on the left-hand side of Eq. 6.20 are neglected, the linear kinematic wave model is obtained. Expressing the second and the third second-order terms in terms of the first on the basis of the linear kinematic wave approximation leads to a parabolic equation (Dooge 1973), in contrast with