Cambridge University Press
978-1-108-49464-9 — Systems of Frequency Distributions for Water and Environmental Engineering
Vijay P. Singh , Lan Zhang
Excerpt
More Information

1

Introduction

1.1 Random Variables in Environmental and Water Engineering

A multitude of hydrometeorologic, hydrologic, geohydrologic, hydraulic, and environmental processes are either random or entail random components that are characterized by random variables. These variables are described by frequency distributions that encompass a broad range. Now the processes and their random components are briefly discussed.

1.1.1 Rainfall

Rainfall is perhaps the most important component of the hydrologic cycle and constitutes input to hydrologic models. The rainfall process is governed by atmospheric processes that result from the interactions between atmosphere, hydrosphere, and the Earth system. The Earth system encompasses the land surface, pedosphere, lithosphere, cryosphere, and anthropogenic influences. The rainfall process is highly heterogeneous in space and time and is characterized, from a hydrologic viewpoint, by intensity, duration, depth of rainfall at a point or over an area; time interval between rainfall events; number of rainfall events in a given time, say, a month, season, or year; extreme rainfall in a year; and areal coverage. These characteristics are random in nature and are therefore described by frequency distributions.

Rainfall depth is often described by an exponential or Weibull distribution. Rainfall duration may be described by an exponential distribution or the distributions skewed to the right. It may be noted that rainfall intensity is given by rainfall depth divided by duration so it may also be described by the same frequency distribution as is for rainfall depth. The number of rainfall events is described by the Poisson or gamma distribution. Extreme rainfall is described by the lognormal, Pearson type III, extreme value type I, or generalized extreme value distribution.

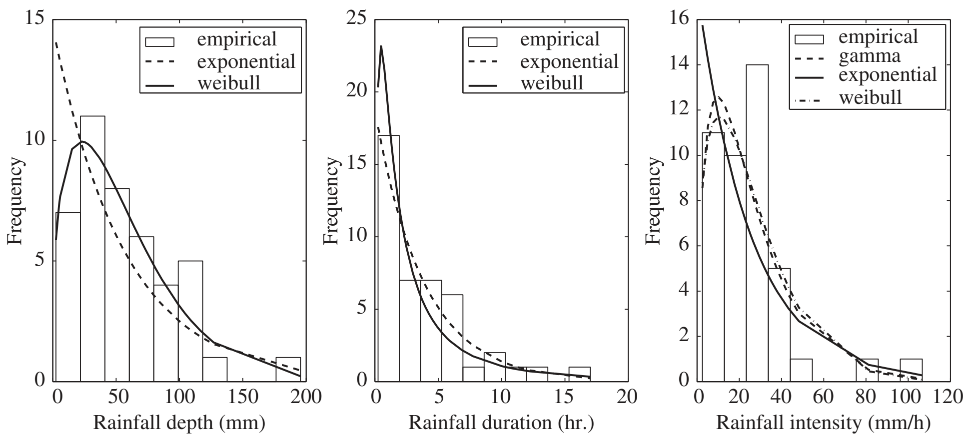

Figure 1.1 Histograms and probability density function plots for rainfall variables, including rainfall depth, duration, and intensity.

Using rainfall depth, rainfall duration and rainfall intensity sample data, Figure 1.1 plots the histogram and the probability density functions estimated from different distribution candidates. It is seen that one may choose different probability distribution functions to model the same set of rainfall variables.

1.1.2 Temperature

Although temperature is a direct component of the hydrologic cycle, it has a decided impact on evapotranspiration, soil moisture depletion, groundwater recharge, vegetation and crop growth, snow and ice melt, freezing and thawing, bacterial and viral growth and spread, and human and ecosystem health. Although temperature field is relatively stable, its maximum and minimum values in a given time, say, month or year; number of days with maximum temperature or maximum temperature; and the starting day of the maximum or minimum temperature are random variables. The maximum temperature is often assumed to follow the generalized extreme value distribution; the minimum temperature is described by extreme value type III or Weibull distribution. Using the monthly temperature as an example, Figure 1.2 graphs the frequency histogram, lognormal, and normal density functions. The frequency histogram indicates two modes. The lognormal and normal distribution might be considered as the two possible candidates without taking the mixture distribution into consideration. In addition, using the number of warm days (i.e., the days warmer than normal during any given month) as an example, Figure 1.3 graphs the frequency histogram and fitted generalized extreme value distribution.

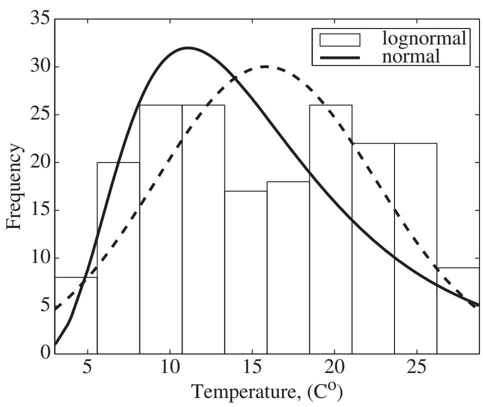

Figure 1.2 Histogram and probability density function plots for monthly temperature.

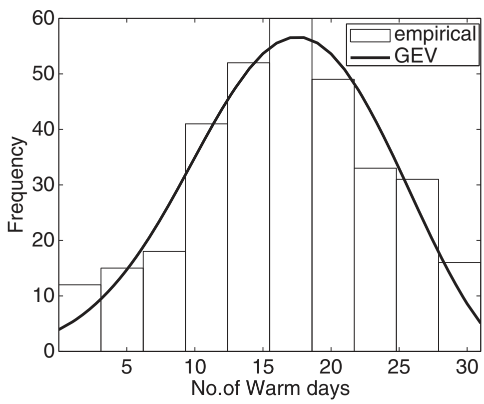

Figure 1.3 Histogram and probability density function plot for number of warm days.

1.1.3 Frost, Fog, and Sunshine Hours

Although sunshine hours at a particular location are relatively stable, they are not the same from year to year. The number of sunshine hours during a given period at a given place can be a random variable and might be described by the normal distribution.

Frost occurs during cold days and has important implications for agriculture. For example, mustard, when it is flowering, is severely impacted by frost persisting for a week or more. The number of frost days (annually) may be considered as random variable and might be described by the normal distribution.

Fog occurs where there is lack of sunshine and has important implications for agriculture, transportation systems, and outdoor recreational systems. The annual number of foggy days may be considered as a random variable and might be described by the Poisson distribution.

1.1.4 Wind

Wind is a regular part of weather. However, extreme winds are becoming too frequent these days. Winds are described by velocity, duration, and direction. Weibull and Rayleigh distribution have been widely applied to model wind velocity (Seguro and Lambert, 2000; Jowder, 2006; Bilir et al., 2015; Pishgar-Komleh et al., 2015; among others). Rayleigh distribution has been found to model wind direction (McWilliams et al., 1979).

1.1.5 Snowfall

In cold regions, snowfall is an important component of the hydrologic cycle. It is a highly heterogeneous process, varying in space and time. The depth of snowfall, duration of snowfall, number of snowfall events in a season, time interval between two snowfall events, and extreme snowfall event in a year are random variables and can be described by frequency distributions. For example, gamma distribution can be used to model extreme snowfall.

1.1.6 Runoff

When it rains, runoff is generated. On urban and small watersheds, overland flow or surface runoff is dominant. Corresponding to rainfall events, runoff events can be defined by their amount, peak value, time to peak, and duration. Although for a given runoff event, these characteristics can be determined deterministically using simple hydrologic models. However, when each characteristic is considered for a duration of time, say a year or several years, then the sample of values of the characteristic can be considered as a random sample and that runoff characteristic might be described by the frequency distributions that are commonly applied to rainfall variables discussed in the preceding text.

1.1.7 Flood

Floods are a common feature in almost every country and cause damage worth billions of dollars, disrupt life and transportation, and cause untold misery almost each year in one part of the world or the other. In hydrology, floods are described by extreme streamflow hydrographs that are characterized by peak, time to peak, volume, duration, and interarrival time. Floods are a random phenomenon when

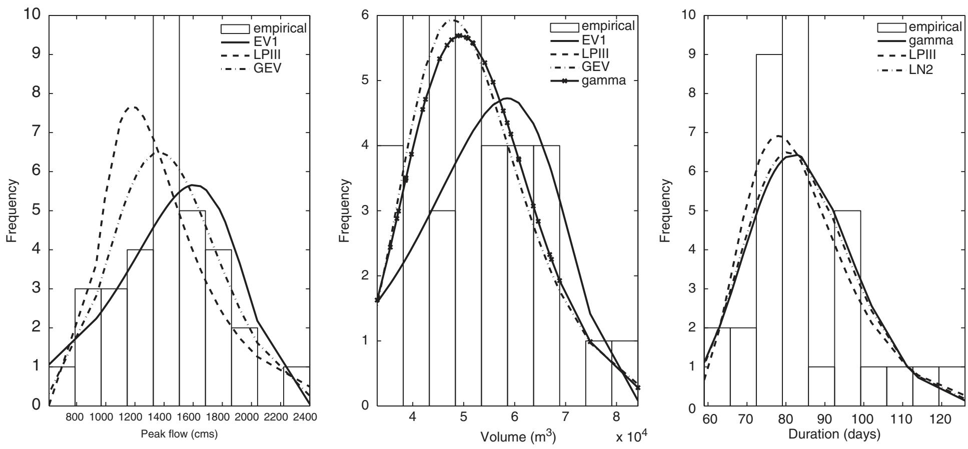

Figure 1.4 Histograms and probability density function plots for flood variables.

maximum from each year or maximum floods above a selected threshold are considered. In that case, the peak value is the instantaneous yearly maximum streamflow. Because of their ubiquitous application to hydraulic design, frequency distributions of floods have been most extensively investigated. Yearly maximum streamflow is often described by the Pearson type III (PIII) distribution in China, log-Pearson type III (LPIII) distribution in the United States, generalized extreme value distribution (GEV) in Europe, extreme value type I (EVI) distribution in certain Asian countries, and 3-parameter log-normal distribution. The interarrival time is described by the exponential distribution, and streamflow volume and duration by the gamma distribution. Figure 1.4 plots the histogram and potential probability distribution candidates for flood variables, including peak flow, flood volume, and flood duration. It is seen that for this dataset, one may choose the GEV distribution to model peak flow. In the case of flood volume, LPIII, EV1, and gamma distributions yield very similar performances. These three distributions (i.e., gamma, LPIII, and LN2) yield very similar performances for flood duration.

1.1.8 Drought

Drought is a creeping phenomenon and impacts all walks of life. Each year, one or the other country is severely impacted by drought, and losses caused as a result are in billions of dollars. There are different types of droughts, such as hydrometeorologic, hydrologic, agricultural, groundwater, and socioeconomic. Hydrometeorologic drought is characterized by precipitation deficit, hydrologic drought by streamflow deficit, agricultural drought by soil moisture deficit, groundwater drought by lowering of water table, and socioeconomic drought by social and financial distress.

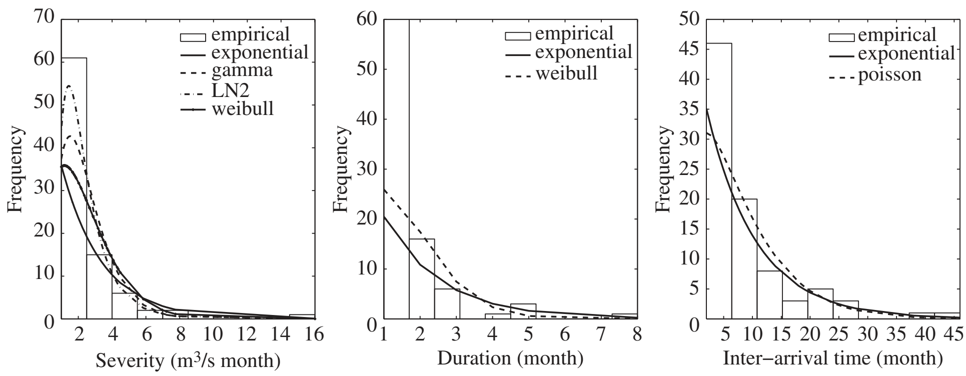

Figure 1.5 Histograms and probability density function plots for drought variables.

Unlike flood, there is no universal definition of drought. For example, drought severity has a relative meaning that may vary from place to place. Nevertheless, regardless of type, drought can be characterized by severity, duration, spatial extent, and interarrival time. All these characteristics are random variables and can be described by frequency distributions. Examples are distributions skewed to the right. Figure 1.5 plots the frequency histogram and probability density functions for drought variables, including drought severity, drought duration, and interarrival time. It is seen that drought severity may be modeled with one of the four distributions, including exponential, gamma, lognormal, and Weibull distributions; Weibull and exponential distributions yield similar performances to model drought duration; and both exponential and Poisson distributions may applied to model drought interarrival time.

1.1.9 Hydrogeology

Hydrogeologic processes are considered as deterministic, for they follow the laws of geophysics. However, there are certain characteristics that may exhibit randomness and may therefore be described by frequency distributions. For example, the distribution of hydraulic conductivity of a geologic formation, such as an aquifer, over space can be considered as a random variable and is often described by the lognormal distribution. Likewise, the average thickness of aquifers can be considered as a random variable that may follow the normal distribution.

1.1.10 Water Quality

Water quality is of fundamental importance in environmental and water engineering and is characterized by a number of parameters. Although these parameters are determined deterministically, they constitute random samples when assembled

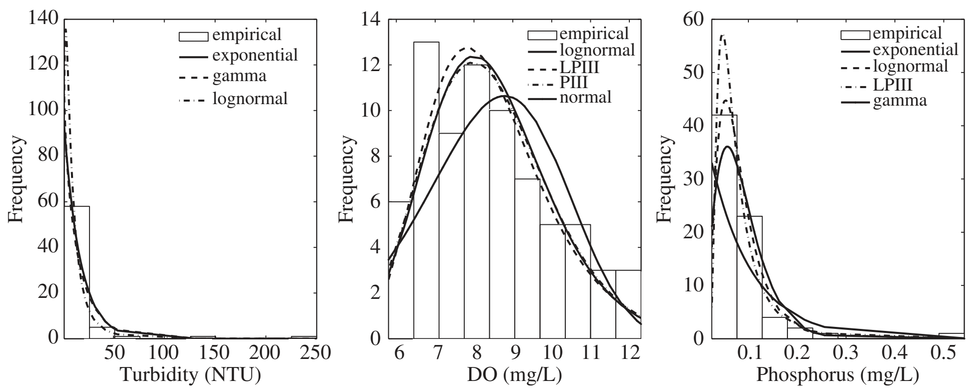

Figure 1.6 Histograms and probability density plots for water quality variables.

over a period, say a few years, or over space, say along a river reach. The sample values can then be described by frequency distributions. In many cases, monthly or yearly water quality variables are skewed to the right. Figure 1.6 plots the histogram and probability density function candidates for monthly water quality variables, including turbidity, dissolved oxygen (DO), and phosphorus. It is seen that the frequency distributions of turbidity and phosphorus are clearly skewed to the right. The right skewed exponential, gamma, LPIII, and lognormal distributions may be applied as probability density functions for turbidity and phosphorus. With the unique characteristics DO, lognormal, LPIII, and PIII, distributions yield similar performances and may be applied as probability density functions for the DO variable. Comparison of the histogram and normal density function indicates that the frequency distribution of DO is still right skewed.

1.2 Systems of Frequency Distributions

It is clear from the preceding discussion that a wide spectrum of frequency distributions are used in hydrologic, hydrogeologic, hydraulic, environmental, and water resources engineering. These frequency distributions employ different hypotheses and have different origins or are derived from different generating systems or families. For example, lognormal, gamma, Pearson and log-Pearson type III, extreme value type I, generalized extreme value, Burr XII, log-logistic, and Pareto distributions are used for flood frequency analysis but do not originate from the same systems or families. Likewise, exponential and Weibull distributions are used for rainfall frequency analysis; 2- and 3-parameter log-logistic distributions are used for analysis of streamflow data, precipitation data, and hazard data; Pareto distributions are used for modeling large exceedances; and Weibull distribution are used for frequencies of low flows and reliability analysis belong to different generating systems. There are different systems from which these frequency distributions are derived and there are different systems of classifications of distributions that are briefly outlined here.

1.2.1 Stoppa System

Stoppa (1990) classified frequency distributions into three groups. The first group comprises three distributions, including Champernowne (1952) distribution, Fisk (1961) distribution, and power distribution (Mandelbrot, 1960). These distributions are derived from probabilistic arguments. The second group of distributions includes gamma distribution (Salem and Mount, 1974), beta distribution (Thurow, 1970), generalized beta distribution (McDonald, 1984), and 3-parameter lognormal distribution (Chieppa and Amato, 1981). These distributions are based on the goodness-of-fit to the observed empirical data that is the common practice in hydrology. However, it is not clear what the hypotheses are used for deriving these distributions. The third group consists of the Pearson system (1895), Burr system (1942), Dagum system, Stoppa system (1990), D'Addario system (1949), and Pareto system (1897), which are derived from differential equations.

1.2.2 Dagum System

Dagum (1990, 2006) recognized three generating systems for classifying frequency distributions: (1) Pearson system (1894, 1895), (2) D'Addario (1949) system, and (3) Dagum system (1980a, 1980b, 1983). Dagum (1977) presented a set of logical-empirical postulates for deriving a particular distribution, which include parsimony, interpretation of parameters, efficiency of parameter estimation, model flexibility, and goodness of fit. Ease of computation and algebraic manipulation may also be added to this set.

1.2.3 Johnson System

Johnson (1949) classified distributions into five groups: (1) Pearson system, (2) expansions, (3) transformed distributions, (4) Bessel function distributions, and (5) miscellaneous. Some distributions may belong to more than one category, while some distributions may not belong to any category. Theoretical arguments have been advanced to justify a particular system of distributions.

1.2.4 General Classification

Singh (2018) suggested nine systems of frequency distributions based on different approaches that seem to have been employed: (1) differential equation involving probability density functions, (2) differential equation involving cumulative distribution function, (3) distribution elasticity, (4) generating functions, (5) genetic theory, (6) Bessel functions, (7) expansions, (8) transformations, and (9) entropy maximization. The frequency distributions resulting from these systems include virtually all the distributions that are known and used in hydrometeorology, hydrology, geohydrology, hydraulics, and environmental engineering.

1.3 Need for Systems of Frequency Distributions

A multitude of frequency distributions have been applied in environmental and water engineering. However, it is not always clear in engineering how different distributions have been derived and what their underlying hypotheses are. It will be interesting to understand these hypotheses that may help select an appropriate distribution for a given set of data. Also, the distribution parameters can be estimated in a more meaningful manner. If different systems of deriving the distributions can be established, then it may be plausible to establish their connections. It may then be possible to develop a universal system from which all distributions or at least a great number of them can be derived. Thus, a study of these systems is of both theoretical and practical interest.

1.4 Organization of the Book

Introducing the theme of the book in the introductory chapter, the subject matter of the book is divided into 10 frequency distribution systems and each system is dealt with in a separate chapter. Each system leads to a number of distributions each of which is derived individually. The derivation sheds light on the hypotheses underlying a given frequency distribution.

References

Bilir, L., Imir, M., Devrim, Y., and Albostan, A. (2015). Seasonal and yearly wind speed distribution and wind power density analysis based on Weibull distribution function. *International Journal of Hydrogen Energy* 40, pp. 15301–15310.

Burr, I. W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics* 13, pp. 215–232.

Champernowne, D. (1952). The graduation of income distribution. *Econometrica* 20, pp. 591–625.

Chieppa, M., and Amato, P. (1981). A new estimation procedure for the three-parameter lognormal distribution. In *Statistical Distributions Scientific Work*, edited by C. Tsallis et al., Vol. 5, pp. 133–140.

D'Addario, R. (1949). Richerche sulla curva dei redditi. *Giornale Degli Economisti e Annali di Economia* 8, pp. 91–114.

Dagum, C. (1977). A new model of personal income distribution: Specification and estimation. *Economic Appliquee* 30, no. 3, pp. 413–436.

Dagum, C. (1980a). The generation and distribution of income, the Lorenz curve and Ginni ratio. *Economic Appliquee* 33, pp. 327–367.

Dagum, C. (1980b). Generating systems and properties of income distribution models. *Metron* 38, pp. 3–26.

Dagum, C. (1983). Income distribution models. In *Encyclopedia of Statistical Sciences*, Vol. 4, edited by S. Kotz, N. L. Johnson, and C. Read, pp. 27–34. New York: Wiley.

Dagum, C. (1990). Generation and properties of income distribution functions. In *Income and Wealth Distribution, Inequality and Poverty*, edited by C. Dagum and M. Zenga, pp. 1–17. Berlin: Springer.

Fisk, P. (1961). The graduation of income distributions. *Econometrica* 29, pp. 171–185.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* 36, pp. 147–176.

Jowder, F. (2006). Weibull and Rayleigh distribution functions of wind speeds in Kingdom of Bahrain. *Wind Energy* 30, pp. 439–445.

Mandelbrot, B. (1960). The Pareto-Levy law and the distribution of income. *International Economic Review* 1, pp. 79–106.

McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrics* 52, no. 3, pp. 647–663.

McWilliams, B., Newmann, M. M., and Sprevak, D. (1979). The probability distribution of wind velocity and direction. *Wind Engineering* 3, no. 4, pp. 269–273.

Pareto, V. (1897). *Cours d'Economie Politique*. New edition by G. H. Bousquet and G. Busino (1964). Geneva, Switzerland: Librairie Droz.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society* 184, pp. 71–110.

Pearson, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Transactions of the Royal Society London* A186, pp. 343–415.

Pishgar-Komleth, S. H., Keyhani, A., and Sefeedpari, P. (2015). Wind speed and power density analysis based on Weibull and Rayleigh distributions (a case study: Firouzkooh county of Iran). *Renewable and Sustainable Energy Reviews* 42, pp. 313–322. doi: 10.1016/j.rser.2014.10.028.

Salem, A., and Mount, T. (1974). A convenient descriptive model of income distribution: The gamma density. *Econometrica* 42, pp. 1115–1127.

Seguro, J. V., and Lambert, T. W. (2000). Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. *Journal of Wind Engineering and Industrial Aerodynamics* 85, no. 1, pp. 75–84.

Singh, V. P. (2018). Systems of frequency distributions for water and environmental engineering. *Physica A* 506, pp. 50–74.

Stoppa, G. (1990). A new generating system of income distribution models. *Quarderni di Statistica e Mathematica Applicata alle Science Economico-Sociali* 12, pp. 47–55.

Thurow, L. C. (1970). Analyzing the American income distribution. *American Economic Review* 48, pp. 261–269.