TALL TALES ABOUT TAILS OF HYDROLOGICAL DISTRIBUTIONS. I

By Vít Klemeš[1]

ABSTRACT: This paper and its companion critique the common frequency analysis techniques for hydrological extremes—in particular, the claims that their increasingly refined mathematical structures have increased the accuracy and credibility of the extrapolated upper tails of the fitted distribution models over and above that achieved by the 50-year-old empirical methods. Part 1 compares the common-sense engineering origins of frequency analysis with its present ostensibly ''rigorous theory''; some myths advanced under the banner of the latter are analyzed in greater detail in Part 2.

> ''If you know something, hold that you know it—and if you do not, admit the fact; this is knowledge.''
>
> Confucius

PREFACE

This paper is based on my 1998 Ven Te Chow Award Lecture, which summarized my views on frequency analysis (FA for short) of hydrological extremes, some of which have been expressed on various previous occasions [e.g., Klemeš (1994), and references therein]. The statistical and probabilistic concepts discussed here are elementary and have been known for decades, but their consequences for hydrology and water resources design are seldom seriously discussed, often ignored or evaded, and almost never acted upon. I should therefore like to hope that writing this paper has not been entirely in vain and that it may help others avoid wasting time and effort on idle pursuits. Looking back, I believe that, had I come across a few papers of a similar kind, say, 40 years ago, I myself might have wasted less time in the early stages of my career, and my Ven Te Chow Award citing ''lifetime career achievement'' may have been better deserved.

INTRODUCTION

It is axiomatic that, for a structure or facility to be safe against water-related hazards, its design parameters should reflect the relevant hydrological extremes, in terms of both their magnitude and frequency. In both respects, the basic source of information is a record of observations of the conditions in the past—the historic record. In hydrology, the relation between magnitudes and frequencies within the historic record has traditionally been depicted by the ''duration curve'' obtained—to follow the old ASCE *Hydrology Handbook* (ASCE 1949)—by arranging the observations ''in order of magnitude, beginning with 1 for the biggest.'' Since historic records of reliable and systematic observations are relatively short, data on hydrological extremes are naturally limited, but one thing is fairly certain—namely, that bigger events than those in the observation record may occur in the future.

This raises the problem of extrapolation beyond the range of observations, which, despite its bad reputation in physics and the exact sciences in general, cannot be avoided in engineering, applied sciences, and everyday life. For the hydrological engineer, one of the most vexing problems has always been **how to extrapolate the upper end of a duration curve**. Help in this matter may one day be provided by the science of hydrology, which, presumably, will someday understand the ''hydrological dice'' well enough to be able to assess the probabilities of possible outcomes from its physical dynamics and geometry, without having to wait for the numbers yielded by its actual throws.

More than a century ago, various empirical formulas for ''maximum flow'' started to appear, without relating this flow to a specific frequency. To quantify the latter, statistical methods were introduced around the turn of the century, and it is important to note that they were treated literally as methods, i.e., as convenient tools for processing the data; as a matter of fact, the duration curve itself is a product of a ''statistical method''—ordering—applied in this pragmatic sense.

Allan Hazen was one of the first to use these methods for extrapolation of hydrological duration curves (of reservoir storage, in this case). The first difficulty he encountered was the fact that a duration curve, when properly plotted as a step function, completely fills its whole definition space (the sample size n) so that it cannot be extrapolated. This is illustrated by the solid line in Fig. 1(a). Extrapolation becomes conceptually feasible only when the discrete steps are replaced by points on a continuous scale of relative frequencies between 0 and 1, because only then ''some free space becomes available'' in the plot before the first and after the last point, as shown in Fig. 1(a), where, following Hazen (1914), the point plot has been flipped horizontally. This horizontal flip he may have done just for convenience; the extrapolation was done manually, and it is more natural to extend lines in the direction of writing, i.e., from left to right, than the other way around.

To make a point representation of the duration curve, it was of course necessary to decide where exactly each point should be plotted within the interval corresponding to its particular step. Common sense led Hazen to pick the midpoint, thus establishing what is now known as the ''Hazen plotting position'' for the rth observation, $PP_r = (r - 0.5)/n$, thereby laying ground to what later was to become the ''plotting position problem''—a major problem in its own right in the context of the FA theory (Cunnane 1978).

But Hazen then encountered another difficulty; in his words:

> ''The practical difficulty with the plotting . . . is the great curvature of the lines. . . . This difficulty is so great as to make the method unsatisfactory in most cases; but this has been removed by using paper ruled with lines spaced in accordance with a probability curve, or, as it is otherwise called, the normal law of error. . . . It is seen that the sharp curvature at the ends is entirely eliminated.''

Plotted on paper ruled in this fashion, the point plot of Fig. 1(a) is transformed into one shown in Fig. 1(b).

It is to Hazen's credit that, in spite of his ingenious disposal of the difficulty with the ''sharp curvature'' of the upper end of the plot, he scrupulously avoided extending his smooth

[1]Retired, 3460 Fulton Rd., Victoria B.C., V9C 3N2, Canada; formerly, Chief Sci., Nat. Hydro. Res. Inst., Envir. Canada.

Note. Discussion open until December 1, 2000. Separate discussions should be submitted for the individual papers in this symposium. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on May 24, 1999. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 5, No. 3, July, 2000. ©ASCE, ISSN 1084-0699/00/0003-0227–0231/$8.00 + $.50 per page. Paper No. 21176.

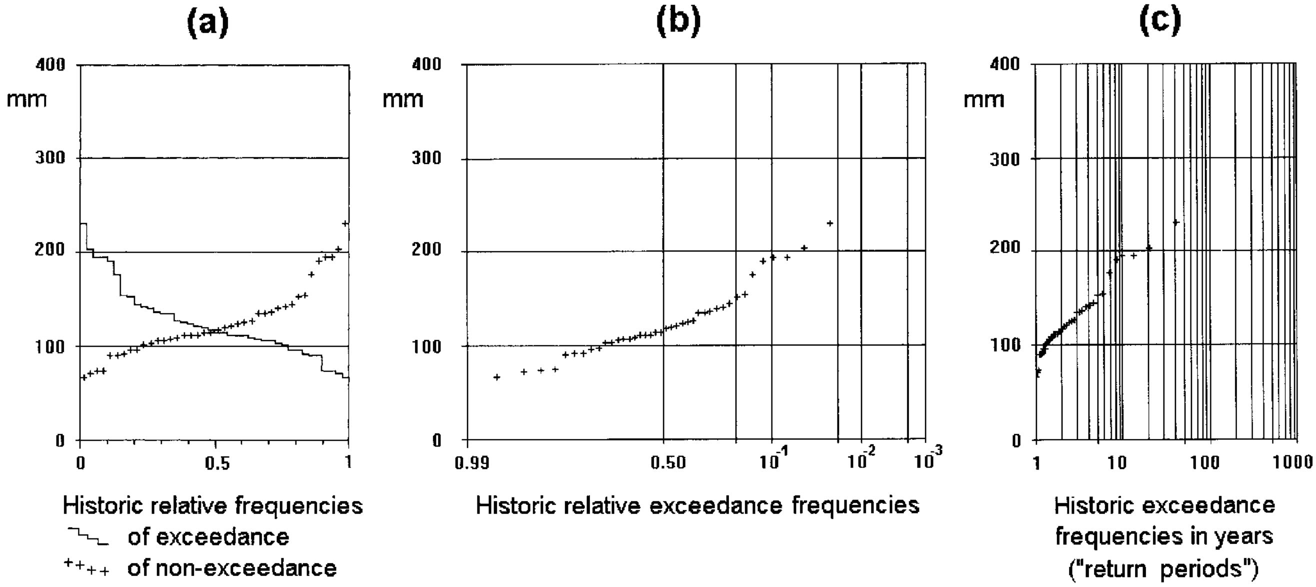

FIG. 1. Different Plots of Historic Frequencies for 40-Year Record of Annual Maxima of Daily Precipitation Totals at Coquitlam Lake, B.C. (a) Two Forms of Duration Curve; (b) Standard EDF on Gaussian Probability Paper; (c) Return Periods on Logarithmic Plot

''lines'' beyond the range 1–99%, cautiously commenting that ''much more numerous data, covering may times longer periods, would be required to settle finally whether the law of error, as used in this way, is strictly applicable to long-term records.''

We thus see that ''statistical methods,'' namely, Hazen's introduction of the ''law of error,'' ''plotting position,'' and ''probability plotting,'' served primarily as a convenience aid in the extrapolation of the upper end of the duration (or frequency) curve. Note that Hazen frankly admitted not knowing the answer to the deeper question, i.e., whether the ''law of error'' had a general applicability to hydrologic data. But his clever invention of the ''probability paper,'' by allowing a geometrically pleasing and practically unlimited extrapolation of any empirical frequency plot, opened the door to speculations whose danger was clearly seen by the great hydrologist of the time, Robert Horton, who warned:

> ''It is, however, important to recognize the nature of the physical processes involved and their limitations in connection with the use of statistical methods . . . [a] Rock Creek cannot produce a Mississippi River flood any more than a barnyard fowl can lay an ostrich egg'' (Horton 1931).

And, fifty years ago, H. K. Barrows, an MIT emeritus professor of hydraulic engineering, did not find the topic of frequency analysis worth more than three pages (including 1 1/2 pages of diagrams) in his 432-page book on floods (Barrows 1948); he described the situation in these simple words:

> ''The most effective method for determining flood frequency consists in plotting frequency in years or percentage of time against peak discharge on logarithmic paper, drawing a smooth curve through the plotted points, and extending this in the higher brackets by eye. Various statistical methods are used in the determination of frequencies. . . . These are merely methods of extending the use of data beyond the length of record; and, since the accuracy of results depends basically upon this length of record, it does not appear that the use of statistical methods adds to the dependability of the results over and beyond that obtained by the use of logarithmic plotting and extension of curves by eye.''

Fig. 1(c) shows such a logarithmic plot for the data from Figs. 1(a and b). Professor Barrows even labeled the frequency axis in his extrapolated plots only as ''probable frequency,'' arguably to stress the lack of any deeper scientific underpinning of the whole exercise. And, demonstrating on several plots how uncertain such extrapolations inevitably are, even for records of the order of 50–100 years, he concluded: ''This shows that [such a] record, or longer than exists on most streams, is not nearly long enough to give valid information on the frequency of a very large flood.''

That is it: Unpleasant and disappointing as it may be for the engineer, the designer, and the planner, the fact that the information they need cannot be provided is freely admitted and clearly stated—no evasions, artful hypotheses, ifs, or let's assumes, and no speculations hidden behind ''scientific'' smoke screens. Confucius would have been pleased with Professor Barrows, Allan Hazen, Robert Horton—he would call them gentlemen—and, indeed, with the intellectual honesty of the scientific culture of the times, still upholding the standards so well epitomized in the famous statement of Newton:

> ''Hitherto I have not been able to discover the causes of those properties of gravity from phenomena, and I feign no hypotheses; for whatever is not deduced from the phenomena is to be called an hypothesis; and hypotheses, whether metaphysical or physical, whether of occult qualities or mechanical, have no place in experimental philosophy'' (Gillott and Kumar 1995).

It is disconcerting to see how this culture has degenerated during the past few decades. Indeed, it seems that 'gentlemen' have been succeeded by 'lesser men,' and Confucius knew the difference: ''A gentleman takes as much trouble to discover what is right as lesser men take to discover what will pay'' (The Analects, Book IV, 16). The latter discovery can be summarized thus: ''If you do not know a thing, redefine it into something you know.''

In the area of hydrological FA, this recipe has been creatively applied as follows: The often intractable real world of large **floods, precipitation, snow accumulations, ice storms**, etc., has been redefined into the well-traveled world of the abstract notion of *random variable*; their **multifaceted** nature has been replaced by a (typically) *univariate* probability distribution; for their **complex** generating mechanisms a *simple* equation with two or three parameters has been substituted and deemed to reign over their whole observed as well as unobserved range; their **historical record** has been proclaimed a *random sample* drawn from the postulated distribution, their **historical frequencies** being thus magically transformed into *time-transcending probabilities*. And there is a bonus: no hydrology, meteorology, climatology, etc., are needed anymore, since all the knowledge required to navigate in this redefined world can be found in some encyclopedia of statistical sciences.

In the process of this ''paradigm shift,'' the concrete practical problem of ''How often can a flood (precipitation, etc.) exceed a specific magnitude at a given location?'' has thus been redefined into a sterile academic exercise on the theme ''What probability distribution model can best be fitted to a given set of numbers?'' And, lo and behold, out of profound ''theoretical analyses'' of the latter, there is supposed magically to emerge an answer to the former—just shake the tail of the resulting ''theoretical probability distribution model'' and all the thousand-, ten thousand-, and million-year events will fall to your feet like ripe pears!

And let's keep in mind that all this is no innocent mathematical parlor game, but a very serious business. To help maintain a real-life perspective, Fig. 1 can serve as a useful benchmark. The data used in it represent a 40-year record of annual maxima of daily precipitation totals for the Coquitlam Lake in British Columbia, and the context for the extent of the desired extrapolation is provided by the fact that the matter of crucial interest here is the exceedance probability of a daily precipitation total of $\approx$400 mm. This value is an estimate of the Probable Maximum Precipitation (PMP), which serves as the basis (via transformation into PMF—the Probable Maximum Flood) for safety assessment of a dam at the outskirts of the city of Vancouver, B.C., Canada (Schaefer 1981; Nikleva 1991).

Professor Barrows would probably have commented that the 40-year record available in the above case ''is not nearly long enough to give valid information on the frequency of a daily precipitation total in that range.'' Indeed, extrapolation of such an extent, no matter what kind of ''tail stretching'' may be used, is a pure guess. But this the modern ''frequency analysis theorist'' cannot admit, since it would immediately expose his ''rigorous'' analyses and estimates as merely tall tales, no better—and arguably worse—than Professor Barrows' ''extension by eye.''

So, let us now peek into the magician's kitchen to see how the tall tales about tails of hydrological distributions are spun.

TWO PILLARS OF THE THEORY OF FREQUENCY ANALYSIS

The two pillars supporting the whole edifice of FA theory are the following hypotheses:

1. A hydrological variable X, such as annual volume of runoff, annual maximum peak flow, or daily precipitation total, is an ''independent identically distributed random variable'' (iidrv) having a (continuous) distribution $F(X)$ of a fairly simple mathematical form.
2. An n-year record of its historic observations is a ''random sample'' from this distribution.

On the one hand, neither of these hypotheses stands on firm hydrologic grounds, nor is their validity a subject of serious research. More typically, statistical hydrology studiously avoids such inquiry (and ignores evidence to the contrary), adopts both assumptions as postulates, and elaborates their consequences by imitating rigorous methods of pure mathematics.

On the other hand, neither of these theoretical pillars was ever needed or had to be invoked in the development of the practical techniques on which FA depends to this day. To the contrary, bringing them on the scene poses some embarrassing problems and requires substantial departures from the very rigor under whose banner they have been introduced, as will be illustrated in the following sections.

The ''iidrv'' and Its Improbable Probability Distribution

As already mentioned, this pillar—the probability distribution of an iidrv—originally was nothing more than a curve intended to smooth out a duration curve to facilitate its extrapolation. The duration curve itself has no inherent probabilistic connotation, since any set of unequal numbers, whether generated by a random or a deterministic mechanism, can be ordered according to their magnitudes. But, during the past three decades or so, we have been exposed to the concept of ''probability distribution of iidrv'' so intensively and persistently that we take its identification with the duration curve for a fact and no longer connect its meaning with hydrological, meteorological, and other pertinent realities—Newton's ''phenomena.''

We know that natural conditions in most basins have been changing over time, that even the climate has been changing and is likely to do so even more, and that there operate large-scale processes (time-wise as well as space-wise) such as ENSO, orbital fluctuations, etc. This all is testimony to nonstationarity and against an ''identity'' of hydrological distributions over time, as well as against sequential independence of hydrological phenomena. We also acknowledge that small and high events may be dominated by different processes, that different mechanisms may prevail at different scales. All this undermines the claim that a given hydrological phenomenon X is identically distributed, i.e., that it is governed by the same simple algebraic form of $F(X)$ over its whole range of magnitude. In short, we know that the whole ''iidrv'' concept is hydrologically impeachable but have been brainwashed into believing that it is ''doing a good job'' (Klemeš 1989).

It is ironic that the only clue the FA theory inadvertently takes from hydrology is the wrong one. It derives the ''distributional assumptions'' [i.e., the general shape of $F(X)$] from a ''probability plot'' such as Fig. 1(b) whose shape is dominated by the small and medium observations. This shape is generally convex on the Gaussian plot, because hydrological phenomena like precipitation, runoff, snow cover, etc., have a zero lower bound, which ''bends'' the lower tail of the plot towards a horizontal asymptote. As a result, all the ''standard'' distribution models are convex on Gaussian frequency scale; they all are models with positive skewness. Hence, **it is the physical regime prevailing in the formation of the lower tail that determines the shape of the extrapolated upper tail**; observations that are hydrologically least relevant to the high extremes—and to the safety of facilities affected by them—have the greatest influence on their estimated ''probabilities''!

Here, the FA theorist faces quite a paradox: On one hand, the observations are supposed to be mutually independent, while, on the other hand, the probabilities of the highest extremes depend on the values of the smallest events in a very deterministic way. To drive this point home, I occasionally lure FA theorists into various traps, one of which is as follows: First, I propose that the magnitude of, say, a 1,000-year flood is unlikely to be affected by conditions in the two or three driest years on record when no floods occurred. This proposition being invariably accepted as reasonable, the following exercise is then performed: (1) some historic record of peak discharges is selected, the analyst selects a distribution model he considers most appropriate, fits it using the most rigorous method available to him, and determines the magnitude of, say, the 1,000-year flow; (2) then the two or three smallest observations are reduced by, say, 50% and the exercise is repeated; (3) finally, the values of the 1,000-year events for cases (1) and (2) are compared. The last time I engaged in this game, the historic flow record was selected randomly by the number of the gauging station, and the results were as follows: the 1,000-year flow was 370 m^3/s in case (1) and something over 700 m^3/s in case (2). Mercifully, when we then checked the name of the gauging station used, we learned it was ''False River.''

One point of such exercise is to demonstrate that the postulate that the flood (or other hydrologic event) is an ''iidrv'' is so strong, it overrides any hydrological consideration, however reasonable or well established it may be. There simply is no way any hydrologically relevant information could ''contaminate'' the sterile Platonic world of contemporary FA. The other point is to show that the unscientific subjective ''extension of the curve by eye'' would be completely unaffected in the above exercise; it would be robust in a hydrologically sensible way.

But even if the ''iidrv'' postulate were justified, it still would be practically useless for adequately specifying the tail of a distribution $F(X)$ in the probability range of interest for design parameters of important structures (0.001 and lower) for purely statistical reasons—because of the shortness of hydrological records. While the Glivenko theorem (Kotz et al. 1983; p. 442) assures ''uniform and almost sure convergence'' of the empirical distribution function, $EDF(X)$, to the true $F(X)$, it would be embarrassingly naive to invoke it here, since the ''sufficiently large'' sample size n to which it refers would have to be of the order of thousands of years to be helpful for the above purpose. This was realized even by mathematically unsophisticated engineers like Hazen and Barrows and was later independently pointed out by the late world-renowned Australian statistician and probabilist, Professor P. A. P. Moran. In the early 1950s, he was engaged in the study of various statistical-hydrological aspects of the Snowy Mountains hydroelectric development, one of them being the estimation of return periods of floods. Professor Moran noted:

> ''This requires, essentially, the estimation of the tail of a probability distribution from a sample of values which is usually not dense in this tail. . . . **These limitations arise from the shortness of the series of observations and cannot be overcome by mathematical sleight of hand. . . . In the first place, the form of the distribution is not known and any distribution used must be guessed**. This may have a considerable effect since the part of the distribution we are interested in is well away from the part where the observations provide some information about the shape'' (Moran 1957; emphasis added).

No doubt, Confucius would have appreciated Professor Moran's sincerity. To my knowledge, the paper in which the above diagnosis was made was Professor Moran's first and last paper on flood probabilities. Seeing that the theory of probability and statistics cannot help engineers in this regard, he considered it useless to engage in any further ''mathematical prestidigitation'' [as he later called it in Moran (1959)] and never touched the subject again. Unfortunately, few FA theorists followed his example. While they sometimes admit that ''it is unlikely that the distribution of [X] can be correctly identified'' (*Estimating* 1988; p. 13), they prefer to emulate Kammerer's and Jung's approach to theories of ESP (extrasensory perception), on which Arthur Koestler has made this cryptic comment:

> ''Like theologians who start from the premise that the mind of God is beyond human understanding and then proceed to explain how the mind of God works, they postulated an a-causal principle, and proceeded to explain it in pseudo-causal terms'' (Koestler 1974; p. 98).

The FA theorists fall into the same trap: They first acknowledge that probabilities of hydrological extremes cannot be determined from the data available and then they proceed to devise ever more rigorous methods for doing exactly that.

The ''Random Sample'' and Its Dilemmas

The cynic might say that the main purpose of the second pillar of the FA theory—the random sample postulate—is to demonstrate that it must be violated in order that any distribution model can be determined in practice. Indeed, the old unscientific approach of extrapolating the duration curve is much less controversial in this regard. It makes only one assumption, which can hardly be questioned—that bigger events can occur in the future than those recorded in the past—and making no others, it violates none either.

Not so with the random sample postulate. First, it assumes that the events are independent, which is admitted to be unlikely even in basic statistical literature: ''Even though our assumption that $X_1, X_2, \ldots, X_n$ are independent is hard to justify in practice, we shall, in the interest of simplicity and an easier exposition, continue to retain it'' (Kotz et al. 1982; p. 607). Second, it implies that the historic record is the result of one of ''repeatable'' and ''equally likely'' experiments, though it is obvious that it is neither, especially given the fact

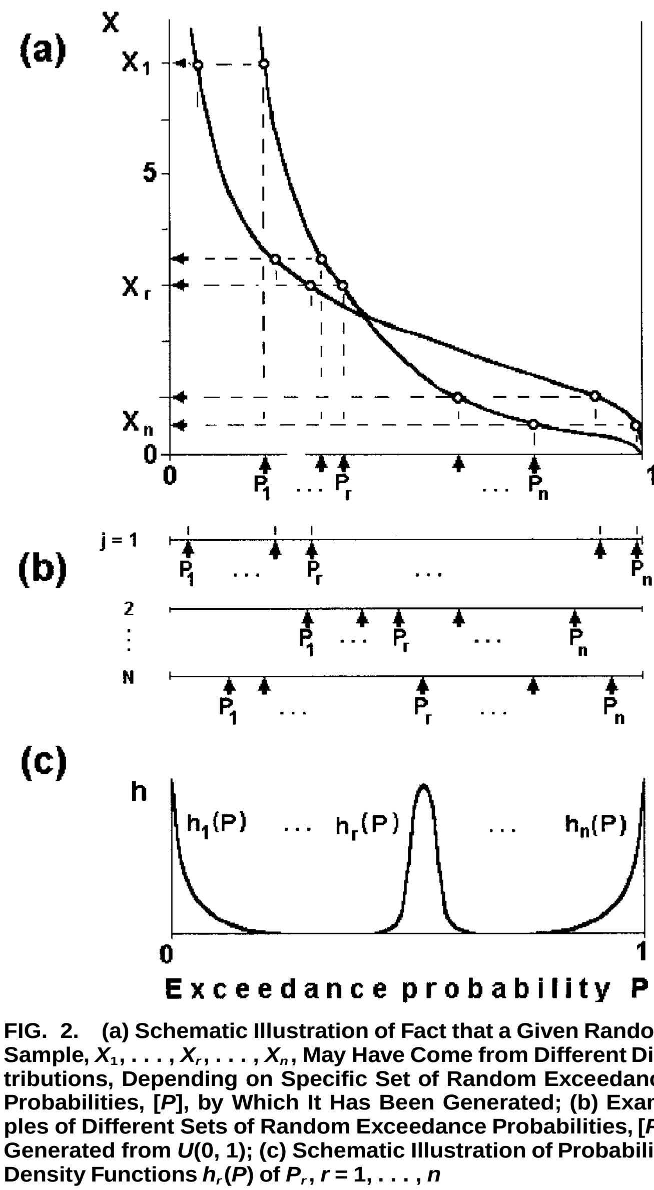

FIG. 2. (a) Schematic Illustration of Fact that a Given Random Sample, $X_1, \ldots, X_r, \ldots, X_n$, May Have Come from Different Distributions, Depending on Specific Set of Random Exceedance Probabilities, [P], by Which It Has Been Generated; (b) Examples of Different Sets of Random Exceedance Probabilities, [P], Generated from $U(0, 1)$; (c) Schematic Illustration of Probability Density Functions $h_r(P)$ of P_r, $r = 1, \ldots, n$

of a nonstationary climate, land use, etc. Third, if observations are ''drawn'' from some specific distribution, then an observation of the same magnitude always has the same probability, no matter which place it occupies in the empirical distribution function (EDF) of a given sample. This is violated by the fact that, in practice, its probability is ''estimated'' by its position, i.e., rank, in the EDF, so that the probability of an observation is not inherent in its own magnitude, as claimed by the theory, but depends on the magnitudes of the other observations in the given sample. Fourth, being ''drawn at random'' from some distribution $F(X)$, the F values (or the exceedance probabilities $P = 1 - F$) corresponding to the random values of X are, by definition, a random sample from a uniform distribution $U(0, 1)$; as such, they are scattered irregularly across the exceedance probability axis and occupy different locations in different samples, so that each P_r itself has a probability distribution $h_r(P)$. This is illustrated in Fig. 2. However, this basic feature of a random sample is violated by the fact that an ordered sample $X_1, X_2, \ldots, X_r, \ldots X_n$, such as shown in Fig. 2(a), is always plotted in regularly spaced plotting positions PP_r, $r = 1, \ldots, n$.

In order to get a better appreciation of the lack of credibility of the extrapolated tails of distribution models supported by the two stated pillars of the FA theory, some of the above points will be examined in more detail in Part 2 of this paper (Klemeš 2000).

APPENDIX. REFERENCES

ASCE Hydrology Committee. (1949). *Hydrology handbook: manuals of engineering practice #28*, ASCE, New York.

Barrows, H. K. (1948). *Floods–their hydrology and control*, McGraw-Hill, New York.

Cunnane, C. (1978). ''Unbiased plotting positions—a review.'' *J. Hydro.*, Amsterdam, 37, 205–222.

Estimating probabilities of extreme floods. (1988). Committee on Techniques for Estimating Probabilities of Extreme Floods, National Academic Press, Washington, D.C.

Gillott, J., and Kumar, M. (1995). *Science and the retreat from reason*, Merlin Press, London.

Hazen, A. (1914). ''Storage to be provided in impounding reservoirs for municipal water supply.'' *Trans. ASCE*, ASCE, New York, 77, 1539–1640.

Horton, R. E. (1931). ''The field, scope and status of the science of hydrology.'' *Trans. Am. Geophys. Union, 12th Annu. Mtg*., National Research Council, Washington. D.C., 189–202.

Klemeš, V. (1989). ''The improbable probabilities of extreme floods and droughts.'' *Hydrology and disasters*, O. Starosolszky and O. M. Melder, eds., James & James, London, 43–51.

Klemeš, V. (1994). ''Statistics and probability: wrong remedies for a confused hydrologic modeller.'' *Statistics for the environment 2 Water related issues*, V. Barnett and K. F. Turkman, eds., Wiley, Chichester, U.K., 345–366.

Klemeš, V. (2000). ''Tall tales about tails of hydrological distributions.'' *J. Hydrologic Engrg*., ASCE, 5(3), 232–239.

Koestler, A. (1974). *The roots of coincidence*, Pan Books, London.

Kotz, S., Johnson, N. L., and Read, C. B., eds. (1982). *Encyclopedia of statistical sciences*, Vol. 2., Wiley, New York.

Kotz, S., Johnson, N. L., and Read, C. B., eds. (1983). *Encyclopedia of statistical sciences*, Vol. 3, Wiley, New York.

Moran, P. M. P. (1957). ''The statistical treatment of flood flows.'' *Trans. Am. Geophys. Union*, 38, 519–523.

Moran, P. M. P. (1959). *The theory of storage*, Methuen, London.

Nikleva, S. (1991). *Revised moisture maximization procedures and PMP values for Coquitlam Lake*, Pacific Meteorology Inc., Richmond, B.C., Canada.

Schaefer, D. G. (1981). ''A study of probable maximum precipitation for the Coquitlam Lake watershed.'' *Rep. No. 14-1474*, BC Hydro, Vancouver, Canada.