TALL TALES ABOUT TAILS OF HYDROLOGICAL DISTRIBUTIONS. II

By Vít Klemeš[1]

ABSTRACT: This paper provides a critical examination of some common ''theoretically based'' approaches to frequency analysis (a general discussion of which appears in Part 1 of the paper) and the myths they have generated about the upper tails of hydrological distributions.

''Believe nothing,
No matter where you read it or who said it,
No matter if I said it,
Unless it agrees with your own reason and common sense.''
Siddhartha Gautama Buddha

INTRODUCTION

As noted in Part 1 of this paper (Klemeš 2000), the theory of hydrological frequency analysis (FA for short) rests on the postulates (1) that the hydrological entity X being analyzed is an ''independent identically distributed random variable'' (iidrv) with a distribution $F(X)$; and (2) that its observation record is a random sample from this distribution.

To paraphrase the late Myron Fiering commenting on spurious mathematization of systems analysis [''We are swept up in a litany of automatic computation, sensitivity analysis, and model making. It has become a new religion'' (Fiering 1976)], one may say that hydrological FA has been swept up in the litany of best fits, efficient estimates, sufficient statistics, unbiased parameters, theorems, and proofs; it has become a new religion, in which the two above postulates play the role of fundamental articles of faith. The following sections will illustrate some of their practical consequences.

THE TAIL WAGGING THE DOG

The methods in which the random sample concept has been used to arrive at a probability distribution model of a hydrological variable fall into two broad categories, which may loosely be labelled ''geometric'' (or graphical) and ''numerical.'' The former relies on a ''best'' fit of the geometry of the duration curve by some analytical curve deemed to represent a ''theoretical distribution model,'' the latter, supposed to be more rigorous, attempting to determine this model via some numerical characteristics of the sample. These were initially based on estimating model parameters by statistical moments, then by the supposedly superior methods of ''maximum likelihood,'' ''maximum entropy,'' and other ''information-theoretical'' methods, whose superiority is now being challenged by the apparently even more superior method of ''L-moments''—which, in a roundabout way, depends on the geometry of the nonrigorous duration curve and its nonrandom features discussed in Part 1 (Klemeš 2000) even more than did the original geometric methods.

To see how this vicious circle closes onto itself despite the ever more clever mathematical sleight of hand deplored already 40 years ago by the late Professor P. A. P. Moran (see Part 1), we shall examine in some detail its first and the (so far) last stages, pausing briefly midway.

Nonrandom Treatment of Random Sample Plot

The original method of estimating the ''parent distribution'' of a hydrological entity was to fit several candidate models to a ''probability plot'' of its observation record (usually with the aid of a linearizing model-specific probability paper) and choose the ''model of the best fit'' on the basis of the minimum sum of (squares of) ''errors.'' It is important to see that what is being treated as ''errors'' and ''minimized'' are the deviations of the observation values X_r from the model values at fixed plotting positions PP_r (e.g., Wallis and Matalas 1974). Thus, in stark contradiction to the underlying assumption that the observations are exact values X taken from an unknown distribution F, it is these ''known-to-be-true'' X values that are treated as ''error corrupted,'' while the representation of their unknown probabilities $F(X)$ by fixed plotting positions is implicitly regarded as error-free, though it is here where the ''errors'' obviously reside. In other words, the observations are treated **not as exact ordinates** of the true distribution function **at unknown and randomly chosen coordinate points** on its F-axis (as the random sample commandment implies and requires), but as **error-corrupted measurements** taken on the fitted curve **at known error-free regularly spaced deterministic coordinates** called plotting positions (PP for brevity).

The ingenious way by which this discrepancy between theory and practice is usually legitimized is by recourse to a result of theoretical statistics which states that, ''given a random sample $X_1, X_2, \ldots, X_n$ from $F(x)$, a **point estimate** of $F(x)$ at an arbitrary but fixed value x is given by $\#(X_k \leq x)/n$, . . . [i.e.,] by the empirical or sample distribution function $EDF(x) = \#(X_k \leq x)/n$'' (Kotz et al. 1985b; p. 320; emphasis added).

The important point here is that all such point estimates, or PPs, at points r {note that the number # defined above is equal to r/n, the familiar ''California PP,'' another point estimate being, for example, the mean of the distribution $h_r(P)$ [see Fig. 2(c) in Part 1], $\langle P_r \rangle = r/(n + 1)$, the well-known ''Weibull PP''}, have this one thing in common: each of them, regardless of its definition equation, **always lies somewhere within the *r*th quantile**. This means that (for the same r and n) a specific point estimate of exceedance probability is always the same, whatever the form of the true parent distribution. For example, the ordered sample $X_n = (X_1, X_2, \ldots, X_n)$ in Fig. 2(a) in Part 1 may have been ''drawn'' from either of the two (or any number of) very different distributions, but, if it were plotted at the same prescribed PPs, the two different distributions would have the same EDF and their ''point'' estimates would be the same for each and every r.

The introduction of the notion of ''point estimate'' may thus mislead one into believing that, even for our embarrassingly small hydrological samples (in contradistinction with the ''sufficiently large samples'' postulated by the Glivenko theorem; see Part 1), it is legitimate to replace the unknown irregularly spaced probability coordinates P_r of the ordered (but random!) observations X_r with the regularly spaced plotting positions PP_r and still get a credible representation of the unknown dis-

[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, June, 2000. ©ASCE, ISSN 1084-0699/00/0003-0232–0239/$8.00 + $.50 per page. Paper No. 21177.

tribution. After all, if we have a valid estimate for each of the n points of the distribution function, does it not mean that the complete set of these estimated points, especially when fitted with a ''theoretical distribution model,'' represents a good approximation of the true distribution, including its upper tail?

The answer is NO. The rank-based ''point estimate'' of the probability P_r is valid only in a similar sense as, say, the **average**, $\langle A\rangle$, of a chronological sequence of random annual flows, $A_1, A_2, \ldots, A_n$, is a valid ''point estimate'' of the flow in an ith year. While this estimate is valid for every single year or point $i = 1, 2, \ldots, n$, nobody would dream of representing the **true pattern of the given annual flow series** by a sequence of these n valid equal point estimates $\langle A\rangle$. But the FA theorist is doing a similar thing when regarding the points plotted at the n **average locations** $r/(n + 1)$ (or quantiles r/n, or other PPs) as credibly representing the **true pattern of the distribution in question**.

The probability that, in a single experiment, each of n random numbers drawn from $U(0, 1)$ falls into its nominal ($1/n$ wide) quantile is very small indeed. Its exact value can be obtained via combinatorics (see Appendix I), but any standard ''table of random numbers'' will serve to provide a good idea; for example, it is virtually impossible to find there a string of ten numbers, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 (in whatever permutation), its exact probability being 0.000363.

In our context, the irregularity of points P_r is most important in the uppermost part of the ordered sample, since in this region their replacement with a regularly-spaced set of PP_r has the greatest effect on the shape of the extrapolated upper tail. This will be examined in detail in the next section.

The Enigma of the Tails

The conventional plot of EDF stretches out, or clips off, the (unknown) distribution tail so that it always extends exactly to the first plotting position, PP_1, on the axis of exceedance probability. The question is: What is the likelihood of such an ''average tail'' occurring in a random sample, and how does it change with sample size? The answer is that this likelihood is small and does not much change with n. In fact, counterintuitive as it may seem, if the observation record really is a random sample, then the **uncertainty in the location of the first (as well as the last) few points of the EDF slightly increases with an increasing sample size**.

This will be demonstrated by a closer look at the distribution of the exceedance probability P_1 of the largest value X_1 in a random sample of size n. Since the exceedance probabilities P_r, $r = 1, 2, \ldots, n$ represent an ordered random sample from $U(0, 1)$, the theory of order statistics defines the density $h_r(P)$ and the distribution function $H_r(P)$ of the distribution of P_r in this form (Kotz et al. 1985b; p. 504):

$$h_r(P) = n \binom{n-1}{r-1} (1 - P)^{n-r} P^{r-1} \quad (1)$$

$$H_r(P) = 1 - (1 - P)^n - \binom{n}{1} (1 - P)^{n-1} P - \binom{n}{2} (1 - P)^{n-2} P^2 - \cdots - \binom{n}{r-1} (1 - P)^{n-r+1} P^{r-1} \quad (2)$$

It may be noted that the use of this distribution for reliability assessment of the exceedance probability (or return period) of the largest observation is not new (Kritskii and Menkel 1981; Lloyd 1995).

For our purpose, we shall use (2), which yields the distribution function of P_1 as

$$H_1(P) = 1 - (1 - P)^n \quad (3)$$

From this, we shall calculate the values of H_1 for the end points of the rth quantiles of a sample of size n as

$$H_1(P = r/n) = 1 - (1 - r/n)^n \quad (4)$$

whose limit for $n \to \infty$ is

$$\lim_{n\to\infty} H_1(P = r/n) = 1 - 1/e^r \quad (5)$$

The quantile end points, r/n, have been chosen as benchmarks because they represent the limiting values of plotting positions (the ''California PPs'').

From (4) and (5) we find that the probability $H_1(P = 1/n)$ of the largest observation X_1 (from a random n-sample) being drawn from the first quantile Q_1 (within which its plotting position PP_1 is **always** located, whatever the formula for its computation) in samples of sizes typical of hydrological records is less than about 65% and is actually slightly **decreasing with an increasing sample size**. The exact numbers are given in Table 1, and a graphical representation of the distribution functions of P_1 for several sample sizes is shown in Fig. 1.

Note that, even for an infinitely large sample, there is almost a 37% probability that X_1 is not in its nominal first quantile, about a 13.5% probability that it comes from beyond the sec-

TABLE 1. Probability that Largest Value, X_1, from Random Sample of Size n Is in First Quantile, Q_1, of Width $1/n$ and that Exactly Five Largest Values, $X_1, \ldots, X_5$, from Random Sample of Size n Are within First Five Quantiles, Q_1–Q_5

Sample size n (1)	Probability X_1 is within Q_1 (%) (2)	Probability $X_1, \ldots, X_5$ are within Q_1–Q_5 (%) (3)
0	67.2	100
10	65.1	24.6
30	63.8	19.2
50	63.6	18.5
100	63.4	17.6
$\to\infty$	63.2	17.55

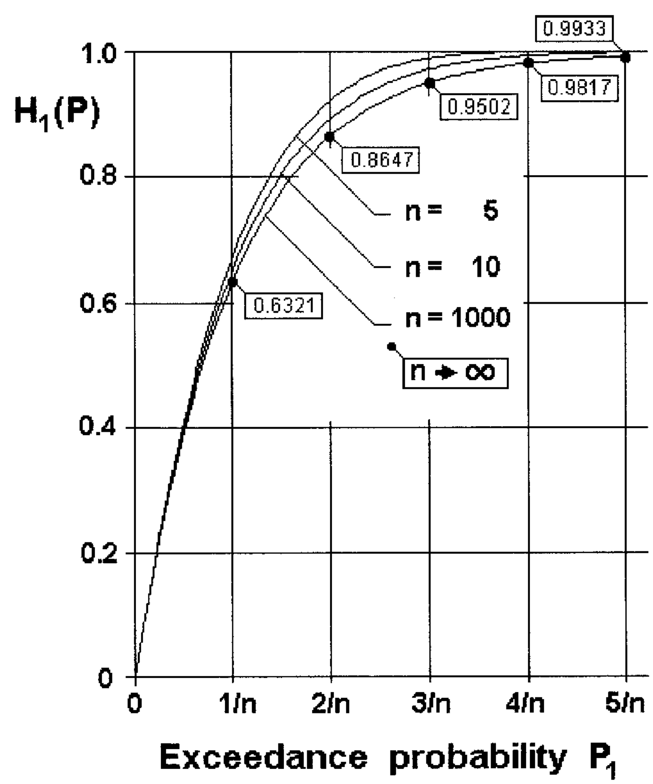

FIG. 1. Distribution Functions of Exceedance Probability, P_1 (Measured in Terms of Quantile Widths $1/n$), of Largest Observation in Samples of Different Sizes n

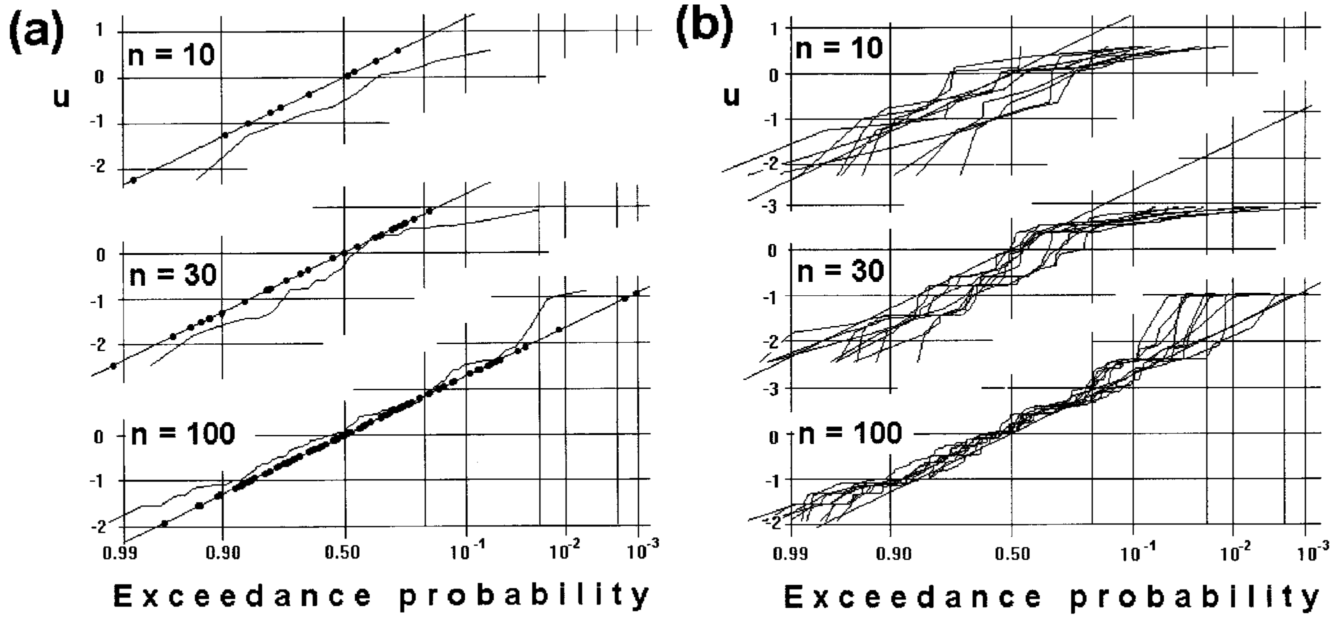

FIG. 2. (a) Typical Random Samples of Sizes n = 10, 30, 100 from Gaussian Distribution; (b) Ten (N = 10) Different "Equally Likely Realizations" of "Probability Plots" of Samples Shown in (a), on Assumption that Their Parent Distributions Are Not Known

ond quantile, about 5% that it comes from beyond the third, and almost 2% that it has been drawn from the fifth or higher quantiles.

On the other hand, a simple computation shows that the probability that exactly the five largest X-values are located in the first five quantiles is generally less than 20% (see Table 1 for exact numbers). However, even in such a case [i.e., **if all the** $P_1, \ldots, P_5$ **were in the interval** $(0, 5/n)$], the probability that each P is located in its nominal quantile is still only 3.84%, while the probability that three or four of the first five quantiles are empty is 9.76%, and **there is a 28.96% probability that three or more of the five smallest P-values are located in the same quantile!** (see Appendix I).

The fact that, **practically regardless of the sample size**, there are such comparatively high probabilities (1) that the largest (as well as the smallest) observations may easily be displaced by up to about five quantiles from their nominal plotting positions; and (2) that they may be dispersed in a very irregular manner, **can distort the shape of $F(X)$ graphically fitted to $EDF(X)$ to such an extent that its extrapolated tail has no greater credibility and objectivity than if the EDF were just "extended by eye."**

The erratic nature of the tails is illustrated in Fig. 2(a), which shows typical random samples of sizes n = 10, n = 30, and n = 100 drawn from a normal distribution, together with conventional plots of their EDFs.

The seriousness of the problem was perhaps best demonstrated by Wallis and Matalas (1974), who found that, for samples of sizes n = 10 to n = 90 from a normal distribution, the minimum-sum-of-squares "best-fits" failed to identify the normal as the parent in ≈40–47% of cases, regardless of the value of n; and when the parent was an extreme value type I (Gumbel) distribution, the same procedure misidentified the parent in ≈60–80% of cases! The significance of this latter result, self-evident as it is, will be brought into even sharper focus in light of the last section of this paper.

So, **if we really do believe** that our observation record X_n was generated as a **random sample** from some distribution, and if we admit that we know neither this distribution, nor the particular set of the ordered random values $P_1, P_2, \ldots, P_n$ which—out of an unlimited number N of such sets that it was capable of "generating"—nature had actually "used" to produce the observation record, **what can we honestly say about the likely form of the distribution $F(X)$?** The answer is: **Little!** Because there is no preferred set of plotting positions for the "probability plot"; our ordered set X_n can legitimately be associated with any ordered set P_n of random numbers from $U(0, 1)$, and each one of them would be an "equally likely realization" of the sample EDF. The best one can do is to generate a number of such "equally likely realizations" to get a better idea of the uncertainty involved. Thus, for example, for each of the three samples shown in Fig. 2(a), ten (N = 10) such realizations are plotted in Fig. 2(b).

Applying this procedure to the real-world sample shown in Fig. 1 of Part 1, one should be making "distributional assumptions" about it not from the EDF in Fig. 1(b) of Part 1—as the common practice would suggest—but from an "uncertainty zone" such as shown here in Fig. 3, which is based on 500 equally likely realizations of the "probability plot" of the data, provided that their parent $F(X)$ is not known.

Needless to say, such an "uncertainty zone" contains very little information useful for making "distributional assumptions" about the form of $F(X)$, especially about the shape of its upper tail. Its uncertainty relates to no hydrological feature but is just a statistical consequence of the specific statistical operation—the ordering—performed on the data, and is the same for any kind of data, given the sample size is the same. In other words, it is merely an empirically constructed confidence band for the plotting positions and can be obtained analytically, for any given confidence level, from (1); for example, Lloyd (1995) calculated the two-sided 95% confidence interval for $PP_{1:50} = PP(X_{1:50}) = 1/51 = 0.0196$ to be (0.0005 − 0.071).

But, paradoxically, this lack of information may offer a better practical guidance for extrapolation of the tail than a "rigorously" fitted and tested "theoretical distribution model" for a number of reasons.

First, regardless of its distorted tails, even the general shape of a standard EDF is misleading for guessing the form of $F(X)$ because of the hydrological reasons discussed in Part 1. Second, a large number of models differing in their upper tails will fit the EDF of a moderately large sample equally well. Third, and perhaps most important, once a specific model is fitted and found to be within some customary (90–95%) confidence band centered on it, by some mysterious psychological process, the "confidence" is automatically transferred from

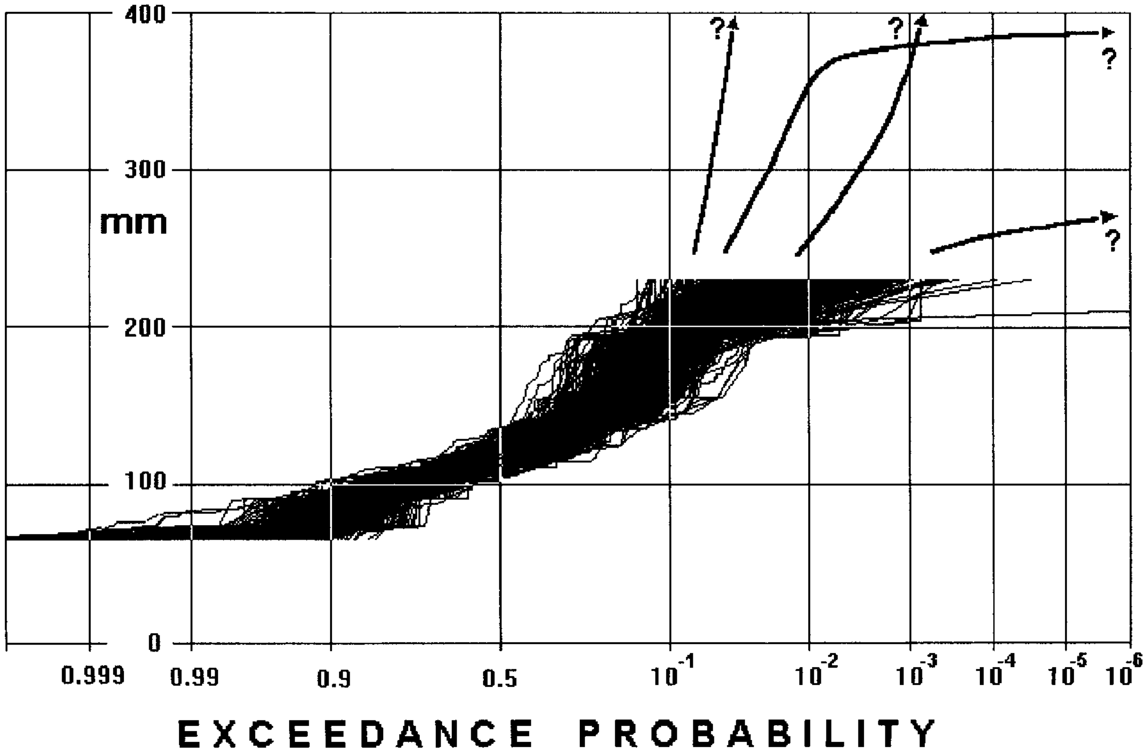

FIG. 3. "Uncertainty Zone" for "Probability Plot" in Fig. 1(b) of Part 1, Based on Its 500 "Equally Likely Realizations," and Some Possible Extrapolations of Its Upper Tail

this band as a whole to the fitted curve itself—the model is no longer treated as merely an ''average'' of the many curves that cannot be ruled out, but, regardless of the most vocal denials, as the one curve that inspires ''confident'' extrapolation. As a distinguished German engineering professor recently complained (Schultz 1993), the idea that, say, the upper 5% confidence limit of the fitted model, rather than the model itself, should be used for the determination of design values ''could not be sold to practice'' during his several decades of experience.

Contrary to the practice outlined above, an ''uncertainty zone'' constructed in the fashion of Fig. 3 makes reference and gives preference to no specific ''theoretical distribution model.'' It therefore does not encourage blind extrapolation of a specific curve from which the desired numbers are just ''read off,'' but makes one think hard, not only as to which way to draw a curve but how far it may make sense to extend it (Fig. 3)—because whoever has ever been faced with such a task ''for real'' (not just as an academic exercise) would know how much more cautious becomes the eye, and heavier the hand, with every centimeter of the extended line. And this is good, because the increasing uncertainty sharpens one's sense of responsibility and encourages one to look for additional sources of information, to consult some real hydrology, meteorology, etc., which the ''rigorous statistical'' approach has shut out (Klemeš 1993, 1996).

JUST A MOMENT!

Statistical moments seemed to offer an ideal escape from the uncomfortable Procrustean bed of geometric distribution fitting. Their computation requires no ordering or other rearrangement of observations, no plotting positions, no distributional assumptions, no curve fitting, no goodness-of-fit testing. Yet they provide information about the basic characteristics of the sample and its distribution: its central tendency, dispersion, symmetry or lack thereof, etc. And if one accepts the two basic FA postulates stated earlier, there seems to be no reason to doubt that the sample moments are approximations of population moments. Better still, since simple distributions (up to three parameter ones) exhibit quite distinct relationships between their moments (''moment ratios''), it seems reasonable to expect that a tendency towards a specific relationship between sample moments will point to a specific distribution; the all-important distributional assumptions will thus be free of subjective elements and follow directly from the observation values themselves.

No wonder, then, that statistical moments have had a great appeal for hydrological frequency analysts who have written extensively about them, including one classical study (Wallis et al. 1974) that bears the suggestive title I have borrowed for this section.

The crucial difference between estimating a distribution model from the moments and from a geometric fit of the EDF resides in this: sample moments may be inaccurate, biased, constrained, etc. (Wallis et al. 1974), but they are not misleading as the shape of the EDF can be because of its artificial regularity forced onto it by the plotting positions. As such, moments involve no ''external contamination'' of the data, they treat each single observation in exactly the same way, and they are as innocent and close to a neutral and objective characterization of reality as one can possibly get—that is, if God, as Einstein believed, while being subtle, is not malicious.

However, Einstein's beliefs notwithstanding, the above straightforward logic breaks down when it comes to inferences that the ''innocent'' statistical moments encourage one to make. For there does not have to be symmetry between top-to-bottom and bottom-to-top inferences, especially in situations involving integration of which moments are an example. The simplest illustration is provided by the arithmetic mean; a given set of numbers gives its value uniquely, but a given mean can result from many completely different sets of numbers. Hence, the sample moment ratios are not of much help. Their diagnostic ambiguity is obvious without any deep analyses and can be seen, for example, from the simple sketch in Fig. 2(a) in Part 1. Its five selected *X*-values certainly do have a unique set of moments and moment ratios, but they could have been drawn from either of the two distributions shown as well as from many others. And this nonuniqueness is not just a consequence of small-sample variability but applies to population moments as well. It is known as the ''moment problem,'' which says, in essence, that it is possible to find two distinct distributions that have the same set of moments (Kotz et al. 1985a; pp. 600–602).

Even when the FA theorist closes one eye and is willing to accept the guidance of the sample moment ratios, he runs into a problem: For different distributions, these ratios have different biases. And, not knowing the parent distribution of the sample, one does not know what biases its moments have, since it could have come from different distributions. So the theorist still has to make a ''distributional assumption'' about the very distribution for which he wants the moment ratios to identify it.

And it is the same story, the same vicious circle, with all the other ''theoretical'' statistical methods of estimation of distribution models: They are overflowing with rigorous analyses of bias, robustness, efficiency, sufficiency, asymptotic convergence, and other exotic properties of models which themselves are the product of manifestly nonrigorous approaches—or guesses, to use the plain word of Professor Moran, as quoted in Part 1.

JUST L-MOMENT!—OR, HAS THE ''PRAYER OF THE STOCHASTIC HYDROLOGIST'' BEEN FINALLY GRANTED?

More than 30 years ago, the late Chester Kisiel proposed the following ''prayer of the stochastic hydrologist'' (Kisiel 1967):

> ''Oh, Lord, please make the world linear and Gaussian!''

While the Lord so far has hesitated, Hosking and Wallis (1997) took up the challenge and, considering their handicap of being mere mortals, have done admirably well: their book has made the world not exactly linear and Gaussian but at least possessing ''linear moments'' (L-moments) and being close to GEV (generalized extreme value) distributed.

L-moments have been a God-send to the FA theorist and practitioner alike. They require no graphical fitting of an EDF plot, and the pages of the above book are overflowing with their other appealing properties. Thus L-moments do not suffer from the ''moment problem,'' but define the distribution uniquely if its mean exists (p. 24); their sample estimates have small biases and near-normal distributions (p. 37); they do not give extreme weight to the extreme observations, as do conventional power moments (p. 38); their moment ratios are robust and not much influenced by outliers and extreme observations (p. 38); they facilitate better specification of, and discrimination between, the underlying population distributions than the conventional moments (p. 40); and they have many other impressive features, including a pedigree of distinguished names, theorems, proofs, and more.

But the most conspicuous practical consequence of their application is that, in regional analyses of annual hydrological maxima, they seem to lead to the GEV distribution much more often than any other method. For this, the most likely explanation is one of these two possibilities: Either most of the worldwide records of annual maxima are indeed samples from GEV distributions (as Gumbel insisted 50 years ago), or such proposition is unsubstantiated (as Moran claimed 10 years later), and the conspicuously frequent tendency to the GEV distribution is an artifact of the L-moment method itself.

The second explanation appears to me more likely than the first for several reasons. To discuss some of them, we need the L-moment definition, which, for the rth L-moment, is given on p. 22 of the above book as follows:

$$\lambda_r = r^{-1} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} E(X_{r-j:r}) \tag{6}$$

where the expectation $E(X_{r:n})$ defined in terms of the exceedance probability $P_{r:n}$ of $X_{r:n}$ has the form

$$E(X_{r:n}) = \int_0^1 x(P) h_{r:n}(P)\, dP \tag{7}$$

with $h_{r:n}(P)$ defined by (1).

The definition offers a first general clue. It suggests why L-moments may make more data sets conform to the same distribution than other models: They involve more levels of averaging. For example, in the geometric fitting, only the probability coordinate of an observation is replaced by an ''average'' (rank, plotting position of some kind), and in conventional moments only the powers of the observations are averaged. As (7) shows, in L-moments one averages functions of order statistics (themselves functions of ''averages'' of exceedance probabilities), then takes linear combinations of these averages as per (6), and, as recommended by Hosking and Wallis (1997, p. 78), in regional analyses one then averages the single-site L-moments so obtained.

I have been skeptical about ''regional averaging'' of distribution parameters for quite some time for a number of reasons (Klemeš 1976), not least because of the fact that real-life structures and facilities are always exposed to the specific local, rather than average, conditions. One has to be cautious in this regard, since not much imagination is needed to envisage that, after one or two more levels of averaging, some ''continental super-moments'' and (Gaussian?) distributions could result, which—in addition to granting the ''stochastic hydrologist's prayer''—could be so ''robust'' that Horton's concern about mixing a Rock Creek with the Mississippi River floods (cited in Part 1) would also be answered: They both would be ''outliers'' and could safely be ignored, since they would have no effect on these super-moments. However, flippancy aside, it is well known that repeated averaging (i.e., integration) reduces the information content of the result, and it should be expected that, after the three levels of averaging involved in ''regional L-moments,'' things tend to become similar no matter what one has started with. A similar equalizing effect of repeated integration has been identified, for example, for cumulative processes (including residual mass-curves of time series): a third-order process of this kind can already be close to an almost-perfect limiting sine wave, whatever the nature of the parent process (Klemeš and Klemeš 1988).

The question now is why this ''equalization process'' seems to favor the GEV distribution. This can be so partly by default, since the locus of L-moment ratios for GEV distribution is close to the middle of the range of loci for the distribution models commonly used. It can, however, be rooted deeper in the L-moment structure. To look into this possibility, we shall examine the following commonly used representations of the first four L-moments (Hosking 1990):

$$\lambda_1 = EX = \int_0^1 x(F)\, dF \tag{8a}$$

$$\lambda_2 = \frac{1}{2} E(X_{2:2} - X_{1:2}) = \int_0^1 x(F)(2F - 1)\, dF \tag{8b}$$

$$\lambda_3 = \frac{1}{3} E(X_{3:3} - 2X_{2:3} + X_{1:3}) = \int_0^1 x(F)(6F^2 - 6F + 1)\, dF \tag{8c}$$

$$\lambda_4 = \frac{1}{4} E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4})$$

$$= \int_0^1 x(F)(20F^3 - 30F^2 + 12F - 1)\, dF \tag{8d}$$

where $F = 1 - P$ is the cumulative, or nonexceedance, probability of X. The polynomials in F on the right-hand sides of (8), designated by the general form $W_r(F)$ and plotted in Fig. 4(a), are simply "weighting functions," or filters, by which the distribution function $F(X)$ [Fig. 4(b)] is successively "processed" to get the values of individual λ_r. In regional analysis comprising N sites, the regional L-moment is obtained as an average, $R\lambda_r = N^{-1} \sum_N \lambda_r$ and, of course, $F(X)$ is replaced with the ordered sample X_n.

The question now can be asked "What kind of a filter, $V(F)$, would produce parameters ξ and $R\xi$ that could be regarded as characteristic of an extreme value population?" One obvious choice for such filter would be one with values $V = 0$ for all the ordered X_r, $r = 1, 2, \ldots, n - 1$, and value $V = 1$ for the sample maximum X_n. In this way it would select only the station extremes, thus producing, by definition, a sample from a population of (regional) extremes which, asymptotically, should have an extreme value distribution. Hence, 100% of the value of the parameters ξ and $R\xi$ constructed with the aid of such a filter could be attributed to an extreme value distribution.

The next question then is "How much (what percentage) of the value of an L-moment λ_r can be attributed to the sample extreme, X_n, or, more generally, to the uppermost part of $F(X)$?" It can be argued that the higher this percentage, the closer the given L-moment is likely to reflect an extremal distribution.

A cursory inspection of the $W(F)$ functions in Fig. 4(a) indicates that the percentage by which the sample extremes contribute to the values of λ_3 and λ_4 could be quite high for distribution functions $F(X)$ with a lower bound equal to zero, small values in the lower tail, and a rather flat body, as illustrated in Fig. 4(b). In such a case, the large weights W for the low F values will be neutralized by the small values of $X(F)$; the effect of the body in the approximate range $0.1 > F > 0.9$ will, to a large extent, cancel out due to the positive-negative symmetry of the $W(F)$ functions in that range, while the high extremes of $F(X)$, roughly in the range $0.9 > F > 1.0$, will dominate, since they combine with the high extremes of all the weight functions $W_r(F)$.

A quantitative illustration of this effect is provided by a sample of size $n = 5$ taken from Fig. 2(a) of Part 1 and represented by order statistics $[X_1, X_2, X_3, X_4, X_5] = [0.5, 1.0, 3.0, 3.5, 7.0]$. The percentages by which the extreme variate X_5 has contributed to the first four L-moments (computed according to the scheme shown in Table 2) are as follows: λ_1 (the mean of X) $\rightarrow$ 47%; $\lambda_2 \rightarrow$ 90%; $\lambda_3 \rightarrow$ 60%; $\lambda_4 \rightarrow$ 61%.

While this is no proof that the geometry of the weighting functions $W_r(F)$ steers regional L-moments toward those of the GEV distribution in the above fashion, the proposition seems worth serious consideration, since some circumstantial evidence suggests that, whatever its exact mechanism, a tendency toward GEV does exist in L-moments. It is known that they tend to identify GEV as the parent even in cases where this is demonstrably not so, for example, where the true parent distribution is bimodal (Gingras and Adamowski 1992). It is interesting in this context to recall that, in complete contrast, the minimum-sum-of-squares fitting of EDFs by Wallis and Matalas (1974) usually could not identify the EV1 distribution as the parent even when this in fact was the case.

As for the various other appealing properties of L-moments cited above, the luster of some of them can, on closer examination, turn out to be an amber traffic light warning of danger. This is especially true about their "robustness" vis-a-vis "outliers," which, in plain language, is just a lack of sensitivity precisely in the part of the distribution which matters most in every safety-related design—in the distribution's high tail. However, this very fact, demonstrated by the virtually unchanged values of L-moment ratios when "dropping the largest observation," has been cited among the strongest arguments why they should be "always preferred in hydrology" (Vogel and Fennessey 1993). What is being praised by these authors is essentially the inability of L-moments to make use of the all-important additional information provided by an actually recorded rare extreme event—a very strange praise indeed in view of the perpetual complaints of hydrologists, engineers, and statisticians alike that it is the paucity of extreme observations that prevents us from making better estimates of

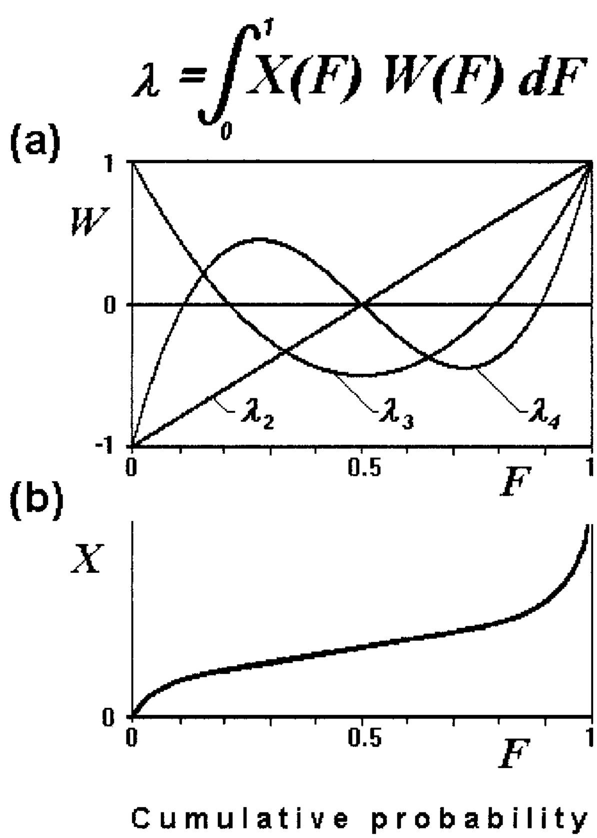

FIG. 4. Illustration of Structure of L-Moments of Distribution $F(X)$

TABLE 2. Computation Scheme for L-Moments of Order j = 2, 3, 4 for Sample Size n = 5 and Observations $X_1 < X_2 < X_3 < X_4 < X_5$

$\binom{n}{j}$ subsamples	$\lambda_{(j=2)}$					$\lambda_{(j=3)}$					$\lambda_{(j=4)}$				
	Weights, w, of ordered observations														
	x_1	x_2	x_3	x_4	x_5	x_1	x_2	x_3	x_4	x_5	x_1	x_2	x_3	x_4	x_5
1	-1	1				1	-2	1			-1	3	-3	1	
2	-1		1			1	-2		1		-1	3	-3		1
3	-1			1		1	-2			1	-1	3		-3	1
4	-1				1	1		-2	1		-1		3	-3	1
5		-1	1			1		-2		1		1	3	-3	1
6		-1		1		1			-2	1					
7		-1			1		1	-2	1						
8			-1	1			1	-2		1					
9			-1		1		1		-2	1					
10				-1	1			1	-2	1					
Σw	-4	-2	0	2	4	6	-3	-6	-3	6	-4	8	0	-8	4
$\binom{n}{j}^{-1} \Sigma w$	-0.4	-0.2	0	0.2	0.4	0.6	-0.3	-0.6	-0.3	0.6	-0.8	1.6	0	-1.6	0.8

the upper tails of hydrologic distributions and better assessments of the safety of our structures!!

For example, in the balmy city of Victoria, B.C., where snow measurements started in 1940 (at the airport) and snow has never been much of a problem, the 57-year maximum of daily snowfall was 34.8 cm and the monthly maximum was 74.7 cm. Then, on December 29, 1996, 64.5 cm of snow fell in one day on top of the 59.4 cm already on the ground, bringing the monthly maximum to 123.9 cm. While the L-moments may have stood up to these ''outliers'' and remained largely intact, large numbers of structures and roofs did not and collapsed. The robustness of the L-moments notwithstanding, most residents reconsidered the robustness of their roofs and snow shovels, and the city authorities recommended, based on the valuable additional information, an increase of snow-load design values.

Ironically, L-moments—the pinnacle of the FA theorist's effort to overcome the inadequacies of distribution fitting by graphical methods on one hand and by conventional moments on the other—may have overcompensated for both. Thus, the artificial ''regularization'' of an EDF achieved in graphical methods by dragging the randomly dispersed observations into evenly spaced plotting positions, instead of being remedied, has been reinforced; the plotting positions were not only brought in through the back door of the ranks of the order statistics, but their role was strengthened by the unequal ''weights'' conferred onto different observations exactly according to these ranks (see Table 2). And the ''defect'' of conventional moments seen in their ''overemphasis'' of high extremes was replaced with a desensitization to these extremes, which, in my judgment, is a defect more dangerous in its practical consequences than their overemphasis. Moreover, this desensitization stands on more shaky theoretical grounds than the ''overemphasis'' of conventional moments, because, in the latter case, the ''overemphasis'' emphasizes nothing but the **actually observed** values, which are multiplied by themselves alone, each being treated the same way as any other. In contradistinction, what is being emphasized in L-moments by the progressively higher powers is the (**unknown!**) coordinates F, which—because they are not known—are replaced with arbitrarily imposed ''average'' coordinates (ranks or plotting positions), the result being an unequal treatment of individual observations, represented by the different weights assigned to them not by nature but by the analyst.

SUMMARY AND CONCLUSIONS

The purpose of this paper has been to argue that the increased mathematization of hydrological frequency analysis over the past 50 years has not increased the validity of estimates of the frequencies of high extremes and thus has not improved our ability to assess the safety of structures whose design characteristics are based on them. The distribution models used now, though disguised in rigorous mathematical garb, are no more, and quite likely less, valid for estimating the probabilities of rare events than were the extensions ''by eye'' of duration curves employed 50 years ago. This is because they rely more heavily on those parts of observation records that may either provide misleading information about the high extremes (the low end of a ''probability plot'') or be largely irrelevant because they can be equally fitted with almost any model (the body of a ''probability plot''). As a result, the bulk of the FA theory, with all its exalted rigor and polish, is spurious, not to say dangerous. It creates an illusion of knowledge where none exists—and illusion of knowledge can do more harm than awareness of ignorance.

Hydrological extremes must, of course, be taken into account when questions of the safety of water-related facilities arise, whether or not their frequencies can be determined scientifically, because, as Ortega y Gasset observed half a century ago, ''Life cannot wait until the sciences have explained the universe scientifically.'' In the meantime, guesses must be made, but it is counterproductive to adorn them with an aura of rigor and science. Rather, in the interest of fair practice, simple extrapolation procedures, commensurate with the current lack of credible scientific basis for extrapolation of distributions' upper tails, should be adopted by professional consensus (Klemeš 1987), and, at the same time, serious work should continue on understanding the ''hydrological dice'' (Eagleson 1972; Klemeš 1978).

Apropos, could it be that Confucius had FA theorists in mind when he said:

> ''Their thinking is insincere because their wishes discolor the facts and determine their conclusions, instead of seeking to extend their knowledge to the utmost by impartially investigating the nature of things.''

APPENDIX I. DISTRIBUTION OF *N* RANDOM NUMBERS FROM *U*(0, 1) OVER SET OF *N* QUANTILES OF EQUAL WIDTHS 1/*n*

The probability of a specific mapping of n random numbers from $U(0, 1)$ onto the set of n equal quantiles $1/n$ can be found by the following combinatorial argument based on Feller (1966).

When a random sample of size n is drawn (with replacement) from a population of n distinct elements, a given element can be drawn k times, $k = 0, 1, 2, \ldots, n$. The actual frequencies with which the n individual elements are drawn are denoted, say, as $x_1, x_2, \ldots, x_n$. If the complete set of frequencies of the k-values in an actual sample is denoted as $y_0, y_1, y_2, \ldots, y_k, \ldots, y_n$ (where y_k is the frequency of k), then the number of different ways to draw an n-sample containing the $n + 1$ possible distinct k-values with their $n + 1$ actual frequencies y_k is given by the formula

$$N = (n!\ n!)/(x_1!\ x_2! \ldots x_n! \quad y_0!\ y_1!\ y_2! \ldots y_n!)$$

where $x_1, x_2, \ldots, x_n \geq 0$; $x_1 + x_2 + \ldots + x_n = n$; and $y_0 + y_1 + y_2 + \ldots + y_n = n$.

Since the number of all possible ways of creating an n-sample from a population of size n is n^n, the probability of a specific kind of sample is N/n^n.

Note that, in the present context, each value x_i represents the number of the actual random exceedance probabilities (i.e., values of P) in quantile Q_i, and y_k is the number of quantiles in which exactly k values of P are found.

ACKNOWLEDGMENTS

I thank Ivo Klemeš of the Department of Mathematics and Statistics, McGill University, for the elaboration of Appendix I. Data on Victoria Airport snowfall were kindly provided by Mr. J. MacDuff, Vancouver Island Special Production Office, Environment Canada, Victoria, B.C., Canada.

APPENDIX II. REFERENCES

Eagleson, P. S. (1972). ''Dynamics of flood frequency.'' *Water Resour. Res.*, 8(4), 878–897.

Feller, W. (1966). *An introduction to probability theory and its applications*, Vol. 1, 2nd Ed., Wiley, New York.

Fiering, M. B. (1976). ''Chapter 17: Reservoir planning and operation.'' *Stochastic approaches to water resources*, H. W. Shen, ed., Colorado State University, Fort Collins, Colo.

Gingras, D., and Adamowski, K. (1992). ''Coupling of nonparametric frequency and L-moment analyses for mixed distribution identification.'' *Water Resour. Bull.*, 28(2), 263–272.

Hosking, J. R. M. (1990). ''L-moments: analysis and estimation of distributions using linear combinations of order statistics.'' *J. Royal Stat. Soc. B*, London, 52(1), 105–124.

Hosking, J. R. M., and Wallis, J. R. (1997). *Regional frequency analysis (an approach based on L-moments)*, Cambridge University Press, Cambridge, Mass.

Kisiel, C. C. (1967). ''Transformation of deterministic and stochastic processes in hydrology.'' *Proc., Int. Hydro. Symp.*, Colorado State University, Fort Collins, Colo., 1, 600–607.

Klemeš, V. (1976). ''Comment on 'Regional skew in search of a parent,' by N. C. Matalas, J. R. Slack, and J. R. Wallis.'' *Water Resour. Res.*, 12(6), 1325–1326.

Klemeš, V. (1978). ''Physically based stochastic hydrologic analysis.'' *Advances in hydroscience*, Vol. 11, Academic Press, New York, 285–355.

Klemeš, V. (1987). ''Hydrological and engineering relevance of flood frequency analysis.'' *Hydrologic frequency modeling*, V. P. Singh, ed., Reidel, Dordrecht, The Netherlands, 1–18.

Klemeš, V. (1993). ''Probability of extreme hydrometeorological events —a different approach.'' *Extreme hydrological events: precipitation, floods and droughts*, Z. V. Kundzewicz, D. Rosbjerg, S. P. Simonovic, and K. Takeuchi, eds., International Association for Hydrological Sciences, Wallingford, U.K., 167–176.

Klemeš, V. (1996). *Estimating the probability of hydro-meteorological extremes (insights from an unorthodox approach)*, B.C. Hydro, Vancouver, Canada.

Klemeš, V. (2000). ''Tall tales about tails of hydrological distributions. I.'' *J. Hydrologic Engrg.*, ASCE, 5(3), 227–231.

Klemeš, V., and Klemeš, I. (1988). ''Cycles in finite samples and cumulative processes of higher orders.'' *Water Resour. Res.*, 24(1), 93–104.

Kotz, S., Johnson, N. L., and Read, C. B., eds. (1985a). *Encyclopedia of statistical sciences*, Vol. 5, Wiley, New York.

Kotz, S., Johnson, N. L., and Read, C. B., eds. (1985b). *Encyclopedia of statistical sciences*, Vol. 6, Wiley, New York.

Kritskii, S. N., and Menkel, M. F. (1981). *Gidrologicheskie osnovy upravleniya rechnym stokom*, Nauka, Moscow (in Russian).

Lloyd, E. H. (1995). *Comments, B.C. Hydro Seminar on Probabilities of Extreme Hydrological Events*, B.C. Hydro, Vancouver, Canada.

Schultz, G. A. (1993). ''Some issues in the application of water resources analysis techniques: what can science sell to practice.'' *Extreme hydrological events: precipitation, floods and droughts*, Z. V. Kundzewicz, D. Rosbjerg, S. P. Simonovic, and K. Takeuchi, eds., International Association for Hydrological Sciences, Wallingford, U.K., 381–392.

Vogel, R. M., and Fennessey, N. M. (1993). ''L-moment diagrams should replace product moment diagrams.'' *Water Resour. Res.*, 29(6), 1745–1752.

Wallis, J. R., and Matalas, N. C. (1974). ''Relative importance of decision variables in flood frequency analysis.'' *Des. of Water Resour. Projects with Inadequate Data; Proc., Madrid Symp.*, UNESCO-WMO-IAHS, 1, 449–456.

Wallis, J. R., Matalas, N. C., and Slack, J. R. (1974). ''Just a moment!'' *Water Resour. Res.*, 10(2), 211–219.