Stochastic Environmental Research and Risk Assessment 14 (2000) 50–68 © Springer-Verlag 2000

L-moments and C-moments

T. J. Ulrych, D. R. Velis, A. D. Woodbury, M. D. Sacchi

Abstract It is well known that the computation of higher order statistics, like skewness and kurtosis, (which we call C-moments) is very dependent on sample size and is highly susceptible to the presence of outliers. To obviate these difficulties, Hosking (1990) has introduced related statistics called L-moments. We have investigated the relationship of these two measures in a number of different ways. Firstly, we show that probability density functions (pdf) that are estimated from L-moments are superior estimates to those obtained using C-moments and the principle of maximum entropy. C-moments computed from these pdf's are not however, contrary to what one may have expected, better estimates than those estimated from sample statistics. L-moment derived distributions for field data examples appear to be more consistent sample to sample than pdf's determined by conventional means. Our observations and conclusions have a significant impact on the use of the conventional maximum entropy procedure which typically uses C-moments from actual data sets to infer probabilities.

1
Introduction

Probability density functions are of central importance in science and engineering, and certainly to us. For example, Mauricio Sacchi (1996) shows how we can build objective functions using Bayes' theorem, Allan Woodbury (1989) is involved in Bayesian interpolation and Tad Ulrych (1998) is examining proba-

T. J. Ulrych
Department of Earth & Ocean Sciences,
University of British Columbia, 2219 Main Mall, Vancouver, BC,
Canada V6T 1Z4

D. R. Velis
Formerly Department of Earth & Ocean Sciences, University of British Columbia; presently Observatorio Astronómico,
Universidad Nacional de La Plata, Paseo del Bosque s/n,
La Plata 1900, Argentina

A. D. Woodbury (✉)
Department of Civil and Geological Engineering,
University of Manitoba, 15 Gillson St., Winnipeg, MB,
Canada R3T 5V6

M. D. Sacchi
Department of Physics, University of Alberta,
Avadh Bhatia Physics Lab., Edmonton, AB,
Canada T6G 2J1

bility density functions associated with chaotic time series. An excellent preprint by Gouveia, de Moraes and Scales (1998) who write about the use of higher moments for the maximum entropy construction of Bayesian a priori pdf's, presents clearly both the use of derived pdf's as well as difficulties encountered in computing these functions from realizations. These difficulties prompted Hosking (1990) to formalize a method for the estimation of statistics that are measures of higher moments, skewness and kurtosis for example, but are more robust with respect to sample size and the presence of outliers. These measures, called L-moments, are based on expectations of linear combinations of order statistics. Indeed, Vogel (1995) observes that the introduction of the theory of L-moments by Hosking "is probably the single most significant recent advance in relating to our understanding of extreme events".

In this paper we concern ourselves, specifically, with the following question. What is the relationship between conventional moments, that we call C-moments in this paper for convenience, and L-moments? This brings into focus such issues as the inversion of L-moments to compute pdf's, the comparison of pdf's computed using C- and L-moments and, finally, the computation of C-moments from L-moment derived distributions.

The robust computation of higher moments of pdf's from limited datasets are often used statistics. Important applications occur in wavelet estimation (Velis and Ulrych, 1996) and in the determination, by means of the principle of maximum entropy, of distributions that enter as prior probabilities in Bayesian inference (Gouveia et al., 1998). The central concept here is that of robustness. The conventional method for the determination of higher order moments, called the direct approach, is by means of an integral formulation. L-moments simplify this computation considerably. Not only do L-moments lead to simplification, they also yield estimates with much lower bias. These properties are of such importance that Royston (1991) has proposed that all statistical packages incorporate L-moments as the method of computing higher order moments. The question remains in our minds, however. How are these different estimates of pdf properties, related?

In order to investigate this topic, we pursue the following methodology. We first compute C- and L-moments from a particular sample size. We then determine the associated pdf using the principle of maximum entropy for both C- and L-moments (no mean task, as it turns out) and compare the derived distributions with the actual pdf. Finally, we compute C-moments from the L-moment derived pdf.

2
Theoretical summary

We present, briefly, the derivation of L-moments, as well as some details of the inverse problem that we solve to determine pdf's from L-moments.

2.1
L-moments

We are all very familiar with the first two moments, the mean and the variance. The next two moments, skewness and kurtosis are also familiar, but less so. In applied geophysics, at least, these moments are seldom used. Although their application has been somewhat limited in the past, the scale dependent phenomena that we are constantly encountered with demands that we look at our data in ever finer detail. This is where these statistics are of great importance. As we will see, conventional estimation of skewness and kurtosis has serious drawbacks. To obviate some of these drawbacks, Hosking (1990) formalized a method for their estimation that is based on expectations of linear combinations of order statistics. This approach originated with the work of Gini (1912) and was described by Kaigh and Driscoll (1987). Hosking's paper, a beautiful paper indeed, has had an explosive effect in some fields. Hoskings's paper has led to several papers that attempt to explain to those of whom, like us, are less well versed with the fine art of pure statistics, what these linear combinations actually mean and what advantageous properties they posses. We are referring to two papers in particular: Royston (1992) and Wang (1996). We present here a distillation that is brief and, we hope, to the point.

2.1.1 A reminder

First of all, a reminder. Consider a random variable X with a pdf $p(x)$. The rth moment of $p(x)$ is defined as

$$E[X^r] = \int x^r p(x) \mathrm{d}x \tag{1}$$

Defining the mean and the variance by μ and σ^2, respectively, we obtain (using Royston's (1992) notation) the usual indices of skewness, $\sqrt{\beta_1}$, and kurtosis, β_2, as

$$\sqrt{\beta_1} = E[(X-\mu)^3/\sigma^3] \tag{2}$$

and

$$\beta_2 = E[(X-\mu)^4/\sigma^4] \tag{3}$$

where $E[\cdot]$ denotes expectation.

2.1.2 Definition of L-moments

Define $X_{j:n}$ to be the jth smallest moment in a sample of size n. The L-moments of X are defined by

$$l_r = \frac{1}{r}\sum_{j=0}^{r-1}(-1)^j \binom{r-1}{j} E[X_{r-j:r}], \quad r = 1, 2, \ldots \tag{4}$$

The first four L-moments are then defined by

$$l_1 = E[X_{1:1}] \tag{5}$$

$$l_2 = \tfrac{1}{2}E[X_{2:2} - X_{1:2}] \tag{6}$$

$$l_3 = \tfrac{1}{3}E[X_{3:3} - 2X_{2:3} + X_{1:3}] \tag{7}$$

$$l_4 = \tfrac{1}{4}E[X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}] \tag{8}$$

Let us first note that we are estimating linear combination of order statistics. In fact the L in L-moments represents exactly this linearity. Wang (1996) gives the following, intuitive, interpretation of L-moments. One value in a sample gives a feel for the magnitude of X. Two samples, through their difference, give a feel for the variability of X. Three samples, give an indication of the asymmetry of

$p(x)$ and, finally, four samples, tell us something about the ratio of peak to tails of $p(x)$. Akin to the definition of conventional normalized moments, $\tau_3 = l_3/l_2$ and $\tau_4 = l_4/l_3$ are statistics related to the skewness and kurtosis of the pdf.

2.1.3
Estimating L-moments by means of PWMs

Greenwood et al. (1979) define PWMs (probability weighting moments) as

$$\beta_r = \int_0^1 x(F)F^r \mathrm{d}F \tag{9}$$

where F is the cumulative distribution.

In a manner similar to the determination of product moments about the mean or about the origin, Hosking has shown that the first four L-moments are given by

$$l_1 = \beta_0 \tag{10}$$

$$l_2 = 2\beta_1 - \beta_0 \tag{11}$$

$$l_3 = 6\beta_2 - 5\beta_1 + \beta_0 \tag{12}$$

$$l_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{13}$$

Given ranked samples of X, $x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n$, Landwehr et al. (1979) have shown that the unbiased estimator of β_r is given by

$$\hat{\beta}_r = \frac{1}{n}\sum_{i-1}^{n} \frac{(i-1)(i-2)\cdots(i-r)}{(n-1)(n-2)\cdots(n-r)} x_i \tag{14}$$

The L-moments may now be computed by means of Eq. (14).

At this stage, Wang (1996) points out that this procedure is not particularly efficient or clear. In other words, PWMs obscure the intuitive understanding of L-moments. Wang makes the point that how L-moments are defined appears to be unrelated to how they may be estimated and he suggests a more logical approach that he calls the direct method.

2.1.4
Direct estimation of L-moments

Consider as a specific example the estimation of l_2, Eq. (6). For each combination of any two values from a population of size n, we form the average of all differences between larger and smaller. Clearly, by examining Eq. (6), we see that this average is simply equal to $2l_2$. Similar considerations apply to the other L-moments. Wang now shows us a method of covering all possible combinations, that may be large even for small sample sizes, to efficiently compute the direct estimators of L-moments.

2.1.5
The Wang scheme

$$\hat{l}_1 = \frac{1}{{}^nC_1}\sum_{i=1}^{n} x_i \tag{15}$$

$$\hat{l}_2 = \frac{1}{2}\frac{1}{{}^nC_2}\sum_{i=1}^{n}({}^{i-1}C_1 - {}^{n-i}C_1)x_i \tag{16}$$

$$\hat{l}_3 = \frac{1}{3}\frac{1}{{}^nC_3}\sum_{i=1}^{n}({}^{i-1}C_2 - 2\,{}^{i-1}C_1\,{}^{n-i}C_1 + {}^{n-i}C_2)x_i \tag{17}$$

$$\hat{l}_4 = \frac{1}{4}\frac{1}{{}^nC_4}\sum_{i=1}^{n}({}^{i-1}C_3 - 3\,{}^{i-1}C_2\,{}^{n-i}C_1 + 3\,{}^{i-1}C_1\,{}^{n-i}C_2 - {}^{n-i}C_3)x_i \tag{18}$$

where, as before, $x_{(i)},\ i = 1, 2, \ldots, n$ are samples ranked in ascending order and

$${}^rC_k = \binom{r}{k} = \frac{r!}{k!(r-k)!} \quad . \tag{19}$$

For a derivation of these equations, please see Wang (1996).

2.2
Maximum entropy density estimation

A very useful, and perhaps the best, method for conservatively assigning probabilities consists of maximizing the entropy of the unknown distribution subject to constraints on its moments (Jaynes, 1957). The problem can be solved by means of Lagrange multipliers in the case that the constraints are given by the conventional moments (i.e. mean, variance, skewness, kurtosis, etc.). Additionally, a normalization constraint is included. The formulation leads to an unconstrained problem where the Lagrange multipliers represent the unknowns.

The entropy of a discrete distribution has been defined by Shannon (1948) as

$$H(p) = -\sum_i p_i \log p_i \quad , \tag{20}$$

where p_i is the probability of one of N possible outcomes of a given experiment. For a continuous distribution,

$$H(p;q) = -\int_a^b p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x \quad , \tag{21}$$

is known as the relative or cross-entropy (Rietsch, 1977; Shore and Johnson, 1981), where x is a continuous variable that lies in the range $[a, b]$, and $q(x)$ is Jaynes' invariant measure. This term represents a state of knowledge against which one makes comparisons. It is also known as the prior.

In practice, a and b are determined from the data (e.g. minimum and maximum values of the sample), and the prior is often the uniform distribution on $[a, b]$ (the least informative prior). A particularly flexible method of incorporating prior information is by means of the principle of minimum relative entropy, that has found important application in the inversion of linear problems of general interest (see Woodbury and Ulrych, 1998 for a complete review).

The normalization is written as

$$\int_a^b p(x)\mathrm{d}x = 1 \quad , \tag{22}$$

and the moment constraints are given by

$$\int_a^b x^r p(x) \mathrm{d}x = \mu_r, \quad r = 1, \ldots, K \ , \tag{23}$$

where μ_r are estimated from the data using sample statistics. Maximizing Eq. (21) with an uniform prior subject to the above constraints, leads to

$$p(x) = \exp\left[-\lambda_0 - \sum_{r=1}^{K} \lambda_r x^r\right] \ . \tag{24}$$

The Lagrange multipliers λ_r are obtained by solving a set of $K+1$ non-linear equations. These equations come from replacing the solution, Eq. (24), into each of the constraints. It is important to note here that the constraints are linear with respect to the unknown distribution. This allowed us to write Eq. (24) as a function of λ_r only.

We propose the use of L-moments instead of C-moments in the maximum entropy determination of probability densities. The advantages of L-moments have already been summarized in a previous section and more details can be found in Hosking (1991) and Royston (1992). Because L-moments are defined in terms of $F(x)$ rather than in terms of $p(x)$, the constraints cannot be easily incorporated into the optimization problem through Lagrange multipliers. Rather, we define the following cost function to be minimized with respect to the unknown distribution:

$$\Phi[p(x); F(x)] = -H + \alpha\left[\left(1 - \int_a^b p(x)\mathrm{d}x\right)^2 + \sum_r (l_r - \hat{l}_r)^2\right] \ , \tag{25}$$

In Eq. (25), the expression between square brackets is a penalty term (α is a constant) that takes into account the constraints. Distribution L-moments, l_r, are evaluated using Eqs. (10)–(13) together with Eq. (9), whereas sample L-moments, $\hat{l}_r$, are computed using Eqs. (15)–(18).

All integrals appearing in the above equations are computed numerically by means of Gauss–Legendre M-point quadrature, where M is the number of points $x_1, x_2, \ldots, x_M$, in (a, b), with which we approximate the distribution, and correspond to the roots of the Gauss–Legendre polynomial (see for example Press et al. (1992)). In practice, we found it very useful to write all unknowns in terms of increments δ_i, that is

$$F_i = F_{i-1} + \delta_i, \quad F_0 = 0, \quad i = 1, \ldots, M \ , \tag{26}$$

and compute p_i using central finite differences:

$$p_i = \begin{cases} (F_2 - F_1)/(x_2 - x_1) = \delta_1/(x_2 - x_1) & i = 1 \\ (F_{i+1} - F_{i-1})/(x_{i+1} - x_{i-1}) = (\delta_i + \delta_{i+1})/(x_{i+1} - x_{i-1}), & i = 2, \ldots, M-1 \\ (F_M - F_{M-1})/(x_M - x_{M-1}) = \delta_M/(x_M - x_{M-1}) & i = M \ , \end{cases} \tag{27}$$

in such a way that the cost function (25) can be expressed in terms of δ_i only. Note that setting $0 \leq \delta_i \leq 1$ is very convenient for ensuring that the F_i's are monotonically increasing and the p_i's are all positive. Then we write

$$H \simeq -\sum_{i=1}^{M} p_i \log p_i w_i \ , \tag{28}$$

and

$$\int_a^b p(x)\mathrm{d}x \simeq \sum_{i=1}^{M} p_i w_i \ . \tag{29}$$

where w_i are the weights of the Gauss–Legendre quadrature. Finally, L-moments are computed through the PWMs using

$$\beta_r = \int_0^1 x(F)F^r \mathrm{d}F = \int_a^b xF^r(x)p(x)\mathrm{d}x \simeq \sum_{i=1}^{M} x_i F_i^r p_i w_i \ . \tag{30}$$

In summary, the optimization problem consists of finding the unknown increment vector $\boldsymbol{\delta}$, such that $\boldsymbol{\Phi}(\boldsymbol{\delta})$ is minimum. Since the cost function $\boldsymbol{\Phi}(\boldsymbol{\delta})$ is highly non-linear, we can either solve the problem by means of a linearization approach, that may depend strongly on a good initial guess, or by using a global optimization algorithm (for example, simulated annealing). Since we believe that the solution of the C-moments problem is, indeed, a good initial guess, we use a local convergent algorithm to minimize the cost function. Here, given $\boldsymbol{\delta}^{\mathbf{j}}$ at iteration j, we compute $\mathbf{F}^{\mathbf{j}}$ and $\mathbf{p}^{\mathbf{j}}$ using Eqs. (26) and (27) respectively. We then compute H and the constraint terms (including the PWMs) as described above.

3
Examples

We now illustrate the behavior of the non-parametric density estimation using the maximum entropy criterion with C- and L-moment constraints. We choose two distributions that reflect only kurtosis and both skewness and kurtosis, and a mixture of two distributions that reflects both skewness and kurtosis, too. Clearly, this is a rather limited investigation, but, we are primarily interested in some broad rather than detailed conclusions. Field data examples are also considered. We examine some realizations of reflection coefficients obtained from well logs, and log-hydraulic conductivity values from an aquifer. All figures also show the pdf derived using a kernel approach, specifically the Epanechnikov kernel (Silverman (1986)), for comparison.

3.1
Symmetric distribution

The first example illustrates the method for samples of various sizes drawn from a Laplace distribution, with and without outliers. The Laplace probability density function with location ξ and scale s is given by

$$p(x) = \frac{1}{2s}\exp(-|x-\xi|/s) \ . \tag{31}$$

This distribution has $\sqrt{\beta_1} = 0$ and $\beta_2 = 24s^2$. For the experiment we set $s = 0.5$ and generated 120 realizations of sample sizes between 10 and 250, and estimated the density function using the first four C-moments and L-moments, respectively. The results of calculating the skewness and kurtosis coefficients from the derived pdf's, as well as the root-mean square (rms) error between true and estimated pdf, are shown in Fig. 1 (1st and 2nd rows). We can observe that the L-moments solution yields smaller rms error for all sample sizes, though the resulting C-moments derived from the density function are no better than the sample C-moments. Although the kurtosis is less biased towards smaller values, its variance is larger than that of the sample kurtosis. A possible explanation for this behavior is that C-moments contain much more information regarding the tail of the distribution than the L-moments. On the contrary, L-moments are insensitive to the tails of the distribution and thus provide more information for modeling the body of the distribution.

Figure 2 compares the resulting pdf for 10 independent realizations of sample size 150. The sharp peak of the Laplace distribution is recovered more accurately using L-moment than C-moment constraints. Table 1 summarizes the mean and standard deviations of the resulting skewness, kurtosis, and rms error for 100 independent realizations of sample size 150.

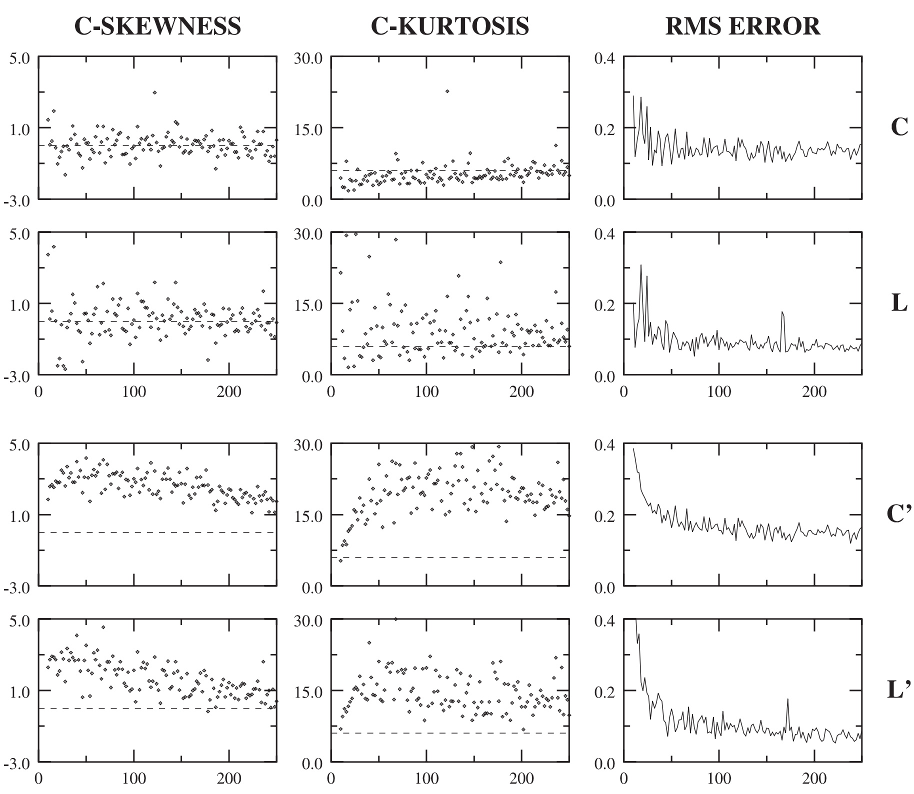

Fig. 1. Skewness, kurtosis and rms error of the estimated pdf's with C- and L-moments constraints for samples drawn from a Laplace distribution without outlier (1st and 2nd rows), and with outlier at $X = 2.5$ (3rd and 4th rows)

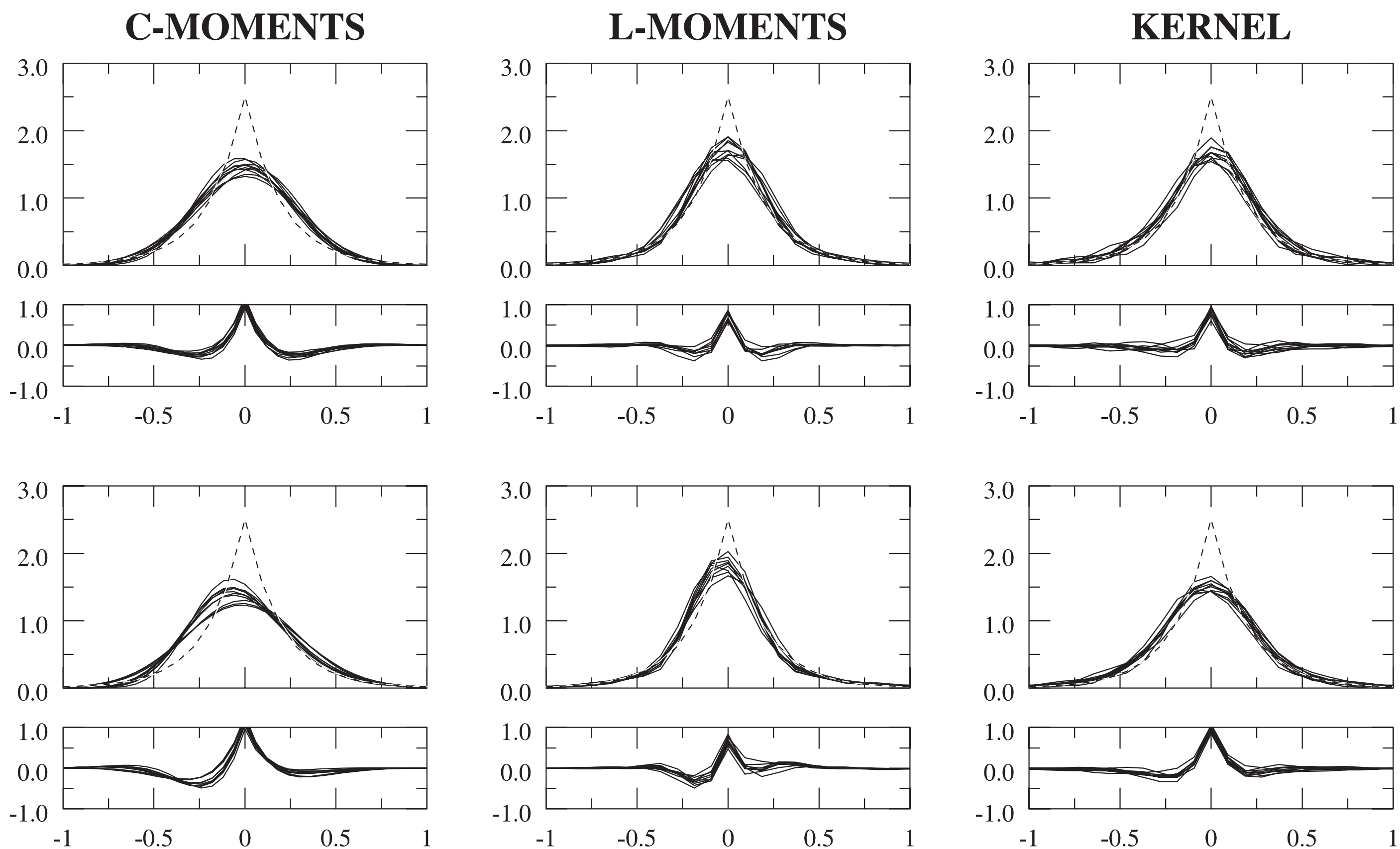

Fig. 2. Estimated pdf's and error for the Laplace distribution without outlier (top row) and with outlier (bottom row). Sample size 150 (10 independent realizations). Dashed line shows the true pdf

Table 1. Mean and standard deviation of the resulting skewness, kurtosis, and rms error for 100 independent realizations of sample size 150 for the Laplace pdf. Primes indicate outlier

	$\sqrt{\beta_1}$	$\sigma_{\sqrt{\beta_1}}$	β_2	σ_{β_2}	rms	σ_{rms}
C	0.02	0.62	5.62	2.23	0.135	0.015
L	0.09	0.76	9.21	4.56	0.084	0.007
C′	2.49	0.48	21.23	3.68	0.153	0.014
L′	1.32	0.69	14.44	4.08	0.088	0.026
true	0.00	–	6.00	–	–	–

The same experiment was repeated but adding a single outlier point at $X = 2.5$. The results are shown in Fig. 1 (3rd and 4th rows). The bias towards higher moment values is now very clear, even for the larger sample sizes. Though C-moments derived from the L-moment constrained pdf have large variance, they are less sensitive to the presence of the outlier. The rms error is also smaller and the shape of the true distribution is more accurately recovered, as may be seen in Fig. 2. Note the failing to reproduce the left shoulder of the pdf in the C-moment constrained case. This is because the right tail is flatter so as to fit the high kurtosis values.

3.2
Skewed distribution

In the next example we used a Log-normal distribution with parameters ξ and s:

$$p(x) = \frac{1}{s\sqrt{2\pi}(x-\xi)} \exp\left[-\log(x-\xi)^2/2s^2\right], \quad x \geq \xi \ . \tag{32}$$

ξ has been selected in such a way that $\mu = \xi + \exp(s^2/2) = 0$ for convenience. For $s = 0.4$, we obtain $\xi = -1.083$, $\sqrt{\beta_1} = 1.32$ and $\beta_2 = 6.24$. The results of the simulations for sample sizes between 10 and 250 are depicted in Fig. 3 (1st and 2nd rows). As observed from the figure, both methods perform well in terms of the resulting C-moments and estimated distributions, though the C-moment constrained pdf error curve shows less variability. Figure 4 shows 10 realizations using sample sizes of 150, and Table 2 summarizes mean and standard deviations for 100 realizations.

The computations were repeated with the addition of a single outlier at $X = 5.0$. As expected the C-moments are heavily biased toward larger values, as illustrated in Fig. 3 (3rd and 4th rows), especially for smaller data samples. The bias is smaller in the L-moment case, while the error is not much different from the no outlier case.

3.3
Bimodal distribution

Perhaps the major impact of using L-moments instead of C-moments can be best appreciated by the following example. Random samples have been drawn from a mixture of two Gaussian distributions $N(-1, 0.2)$ and $N(1.0, 0.2)$, where the proportion of the mixture was 40% and 60%, respectively. We run the same experiment as in the previous examples and the results are displayed in Figs. 5 and 6. Here, we added a single outlier at $X = 7.0$. In all cases the rms error of the L-moment constrained pdf is significantly smaller that the C-moment counterpart. At the same time, the conventional skewness and kurtosis measures

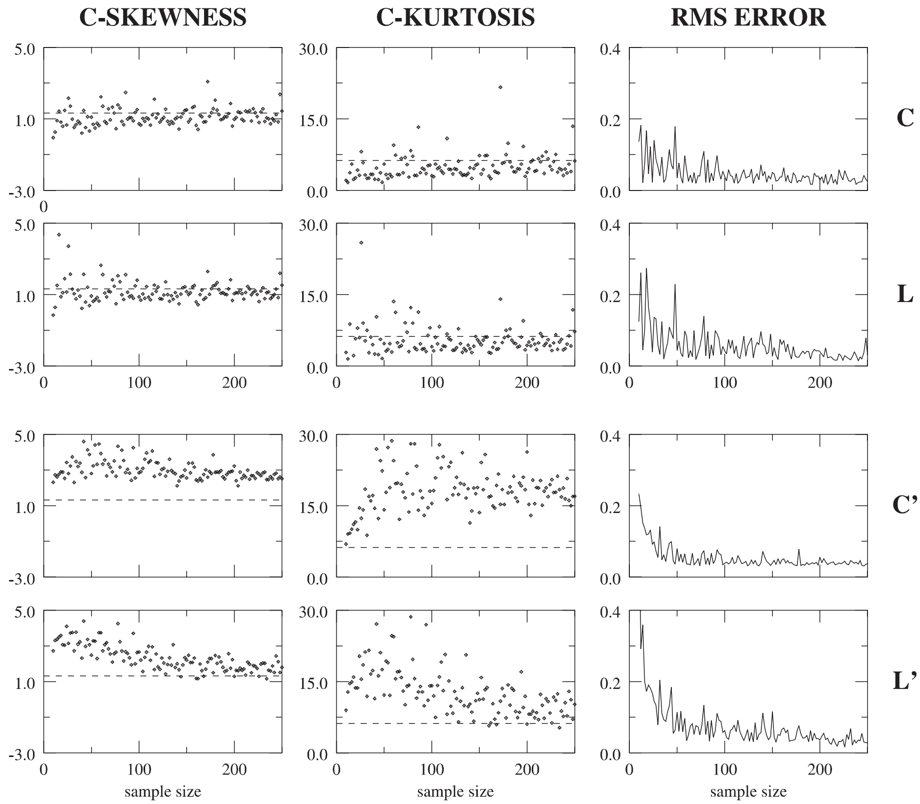

Fig. 3. Skewness, kurtosis and rms error of the estimated pdf's with C- and L-moments constraints for samples drawn from a Log-normal distribution without outlier (1st and 2nd rows), and with outlier at $X = 5.0$ (3rd and 4th rows)

obtained from the resulting L-moment constrained pdf show less sensitivity to sample size and outliers than the sample statistics. In some cases, the two modes of the distribution can only be well-resolved using L-moments, as illustrated in Fig. 6. When the outlier has been added, both the C-moment (constrained) and the kernel pdf's fail to reproduce the true distribution.

Although not shown here, the use of a larger sample size does not affect significantly the estimation of the pdf using C-moments, unless more than the first four moments are utilized. On the other hand, the first four L-moments appear to contain more complete information about the distribution shape than the first four C-moments. In this case, and as opposed to L-moments, C-moments are very sensitive to the tails of the distribution, and not very good at characterizing its center, as shown in Fig. 5 and Table 3.

3.4
Field examples

We obtained interesting results while computing C- and L-moment derived pdf's for some realizations of reflection coefficients obtained from well logs in the Campos basin, offshore Brazil (Rosa and Ulrych, 1991). We examined many different logs but show the results of only one for obvious reasons. These results are, however, remarkably representative. Figure 7 (top row) shows the derived

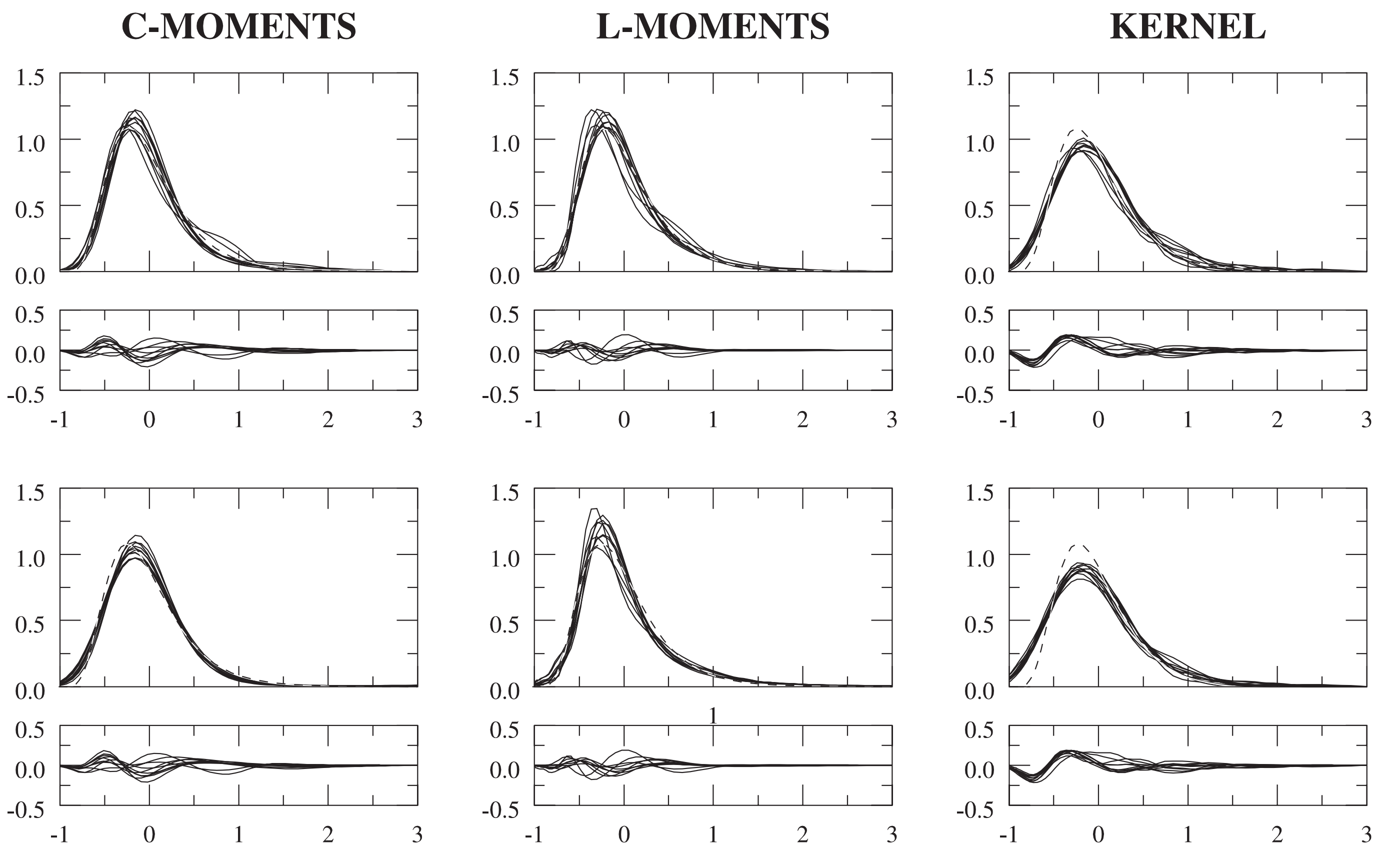

Fig. 4. Estimated pdf's and error for the Log-normal distribution without outlier (top row) and with outlier (bottom row). Sample size 150 (10 independent realizations). Dashed line shows the true pdf

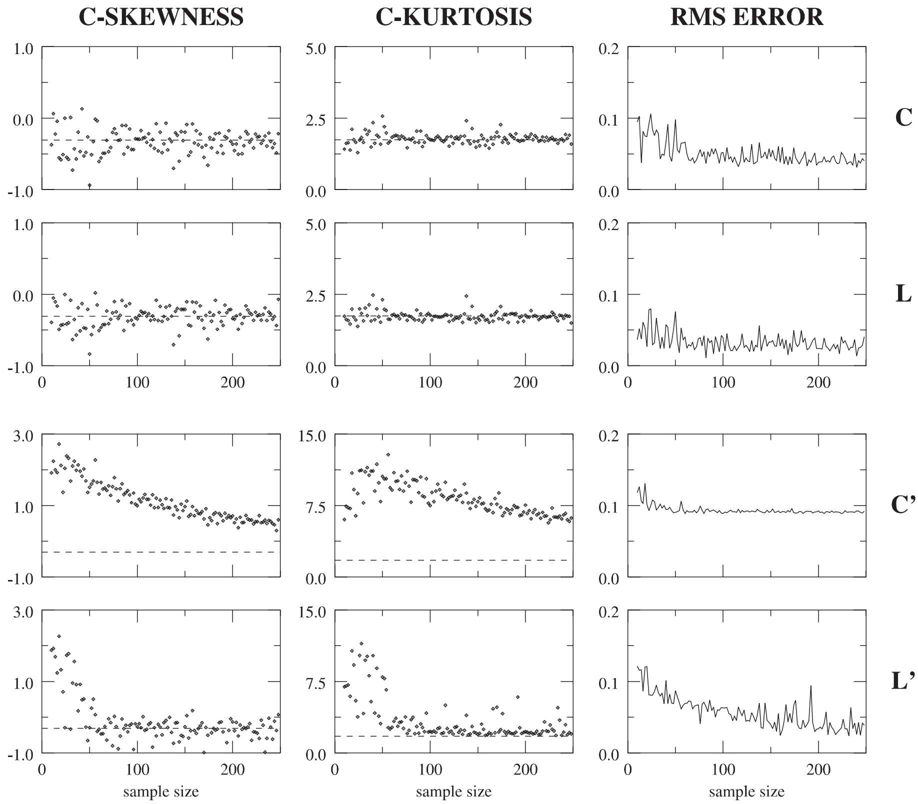

Fig. 5. Skewness, kurtosis and rms error of the estimated pdf’s with C- and L-moments constraints for samples drawn from a mixture distribution without outlier (1st and 2nd rows), and with outlier at $X = 7.0$ (3rd and 4th rows)

pdf’s when the whole sample containing 1000 points is considered. Figure 7 (bottom row) shows the results when the pdf’s are computed with 100 point non-overlapping windows to avoid non-stationary effects in the computation. Apart from a single 100 point realization that yields a broad bell shaped distribution on both the C- and L-moment pdf’s, the L-moment derived pdf’s show more consistency, sample to sample.

A second field example consists of log-hydraulic conductivity values from the Borden aquifer in Ontario, Canada. This particular data set, and subsequent interpretations have sparked considerable controversy in the literature (Sudicky, 1986). Domenico and Schwartz (1998, p. 38) make the point that “there is no hard and fast rule that hydraulic conductivity is log-normally distributed, or for that matter are other parameters”. Indeed, Woodbury and Sudicky (1991) show that other distributions may be appropriate for a valid description of the hydraulic conductivity. Woodbury and Sudicky (1991, Figs. 1, 2) also note the presence of outliers in the data set, i.e. values of ln-K of less than -6.0, and that the data set has a significant negative skew. What is then, the effect of outliers on the parametric modeling of such a data set? In Fig. 8 we present the estimated pdf’s using 100 samples out of the total data set of 1188 values of ln-K data (solid line). The sub-sampling procedure is used to examine statistics from a data set affected by correlation (Woodbury and Sudicky, 1991). In this

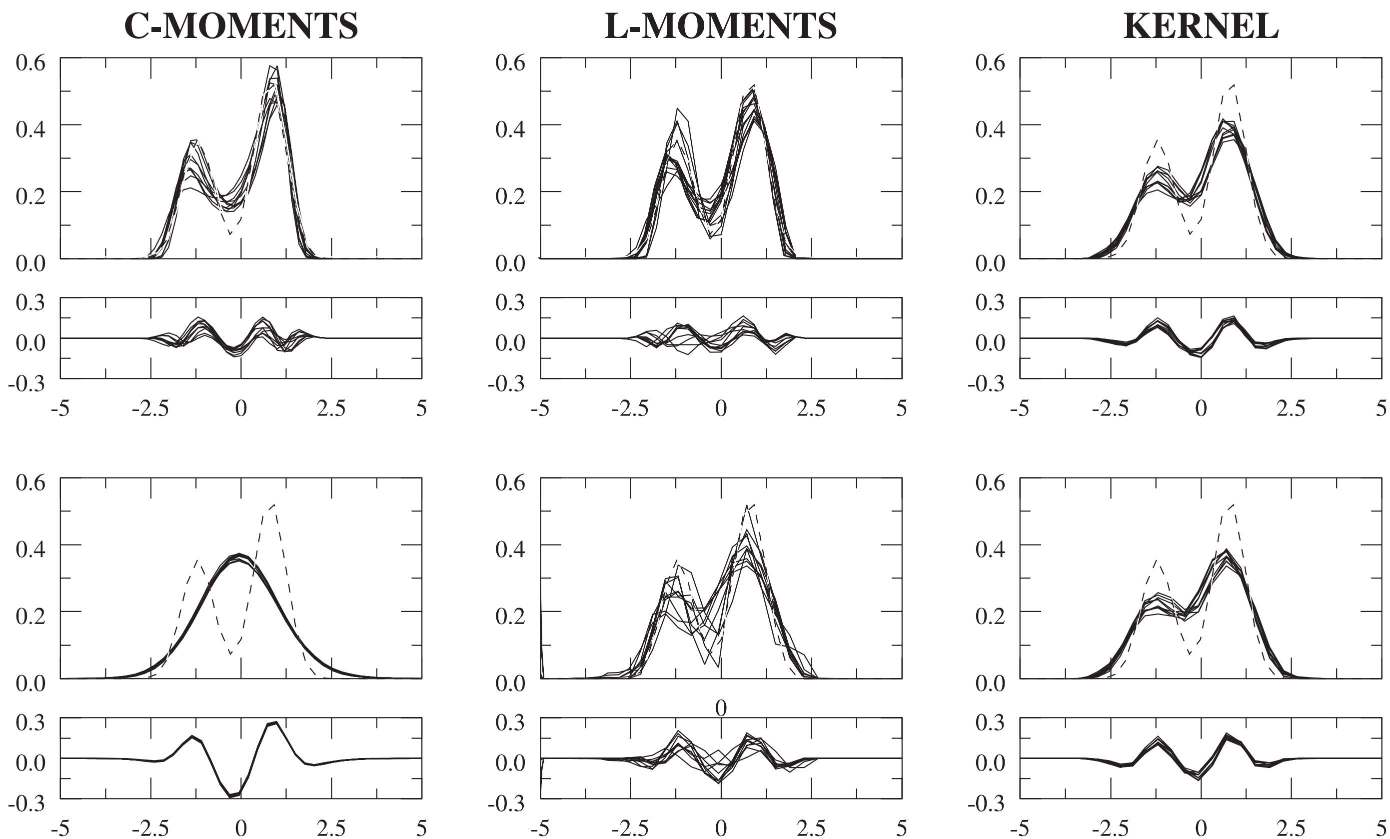

Fig. 6. Estimated pdf's and error for the mixture distribution without outlier (top row) and with outlier (bottom row). Sample size 150 (10 independent realizations). Dashed line shows the true pdf

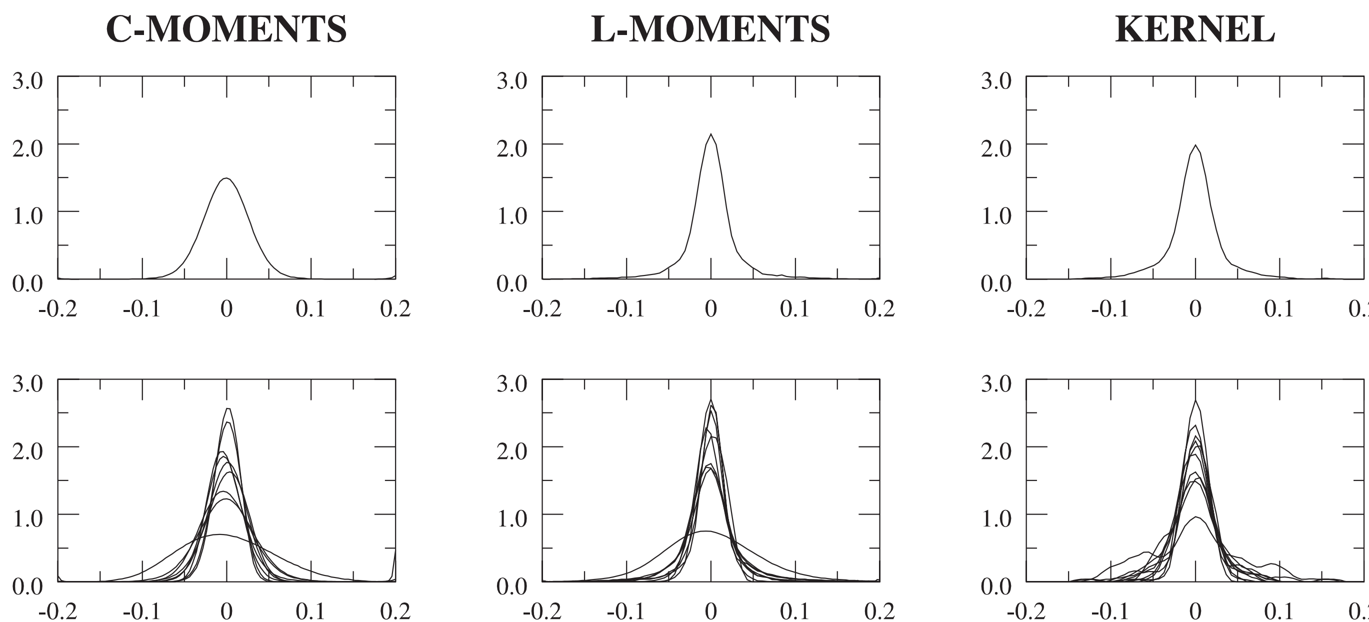

Fig. 7. Reflection coefficient pdf's of a well log in the Campos basin, offshore Brazil, from 1000 points (top row) and 100 point non-overlapping windows (bottom row)

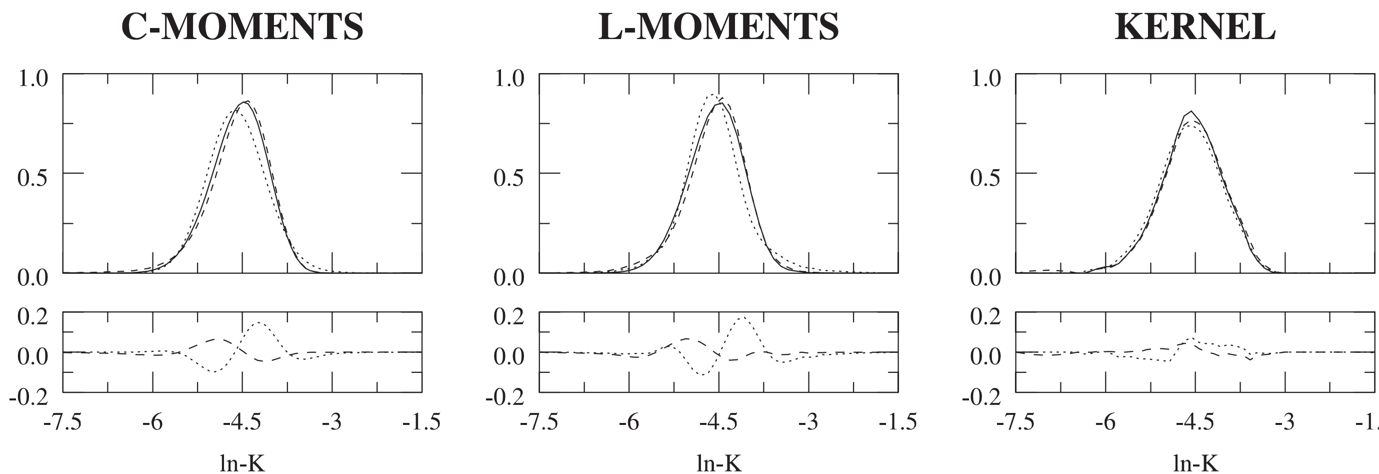

Fig. 8. ln-*K* pdf's. Data from the Borden Aquifer, Ontario Canada (Sudicky, 1986). Smaller plots show the difference between the no-outlier and the outlier cases

Table 2. Mean and standard deviation of the resulting skewness, kurtosis, and rms error for 100 independent realizations of sample size 150 for the Log-normal pdf. Primes indicate outlier

	$\sqrt{\beta_1}$	$\sigma_{\sqrt{\beta_1}}$	β_2	σ_{β_2}	rms	σ_{rms}
C	1.23	0.34	5.33	1.89	0.034	0.013
L	1.21	0.35	5.37	1.95	0.036	0.017
C′	2.92	0.29	18.76	3.35	0.038	0.007
L′	2.11	0.37	11.45	2.91	0.048	0.017
true	1.32	–	6.24	–	–	–

Table 3. Mean and standard deviation of the resulting skewness, kurtosis, and rms error for 100 independent realizations of sample size 150 for the mixture pdf. Primes indicate outlier

	$\sqrt{\beta_1}$	$\sigma_{\sqrt{\beta_1}}$	β_2	σ_{β_2}	rms	σ_{rms}
C	−0.33	0.13	1.79	0.16	0.044	0.015
L	−0.29	0.12	1.76	0.16	0.032	0.008
C′	0.86	0.11	7.79	0.55	0.091	0.001
L′	−0.37	0.27	2.52	0.80	0.047	0.012
true	−0.31	–	1.74	–	–	–

Table 4. Skewness and kurtosis for data from the Borden Aquifer

	no outlier		ln-K = −7.0		ln-K = −1.0	
	C	L	C	L	C	L
$\sqrt{\beta_1}$	−0.31	−0.24	−1.06	−0.70	1.92	0.53
β_2	3.02	3.20	6.27	4.85	14.48	5.83

particular case, there is almost no visual difference between the C- and the L-moment solutions. The calculations were also repeated including a single outlier at ln-$K = -7.0$ (dashed line), and at ln-$K = -1.0$ (dotted line), in order to simulate possible data errors. In the second case (outlier at ln-$k = -1.0$), sample skewness and kurtosis go from -0.31 and 3.02 to 1.92 and 14.48, respectively. On the contrary, skewness and kurtosis derived from the L-moments constrained pdf are 0.53 and 5.83 respectively, values that are much closer to those of the no-outlier case (see Table 4).

4
Summary and conclusions

We have attempted to answer, or at least explore, the question of how C- and L-moments are related. The approach we have used is to compare these moments by means of the pdf's that can be computed by using them as constraints in the inversion. First, we summarize in the form of a partial list, some of the properties of L-moments as reported in the literature.

1. L-moments, as compared to C-moments, are linear functions of the data and suffer less from effects of sampling variability, are more robust to outliers and are less sensitive to sample size (Hosking, 1991). Of particular significance is the much larger bias that $\sqrt{\beta_1}$ and β_2 exhibit for small sample sizes, in comparison to L-moment ratios.
2. In general, Royston (1991) points out that whereas $\sqrt{\beta_1}$ and β_2 are very sensitive to small and perhaps unimportant perturbations in the tails of the distribution, τ_3 and τ_4 are dependent on changes in the shape of the main portion.
3. A Normal distribution, $N(\mu, \sigma^2)$, may be described in terms of its L-moments as $l_1 = \mu$ and $l_2 = \pi^{-1/2}\sigma$. Unfortunately, according to Hosking, an L-moment analog of covariance is not easily defined.
4. Another observation by Hosking that has unfortunate ramifications for the inversion problem (see Gouveia et al., 1998), is that no extension of L-moments to the multivariate case is immediately apparent.
5. Whereas in some cases (Cauchy distribution being an example) C-moments may not exist, L-moments are guaranteed to exist.

The points enumerated above, clearly describe L-moments as statistics that show many advantages in comparison to conventional measures. We have also found, in our simulations, that, in general, L-moments are much less sensitive to sample size and, in particular to outliers. Perhaps the examples that are illustrated in this paper do not underline this conclusion, but certainly the work of other researchers as well as many examples that we do not have room to report, strongly support this observation. Given these properties, we had hoped that pdf's computed using L-moments could then be used to compute C-moments that would be more robust measures than those derived from sample statistics. This has not turned out to be unequivocally so. It is true that the pdf's derived from L-moment constraints are "better" than those derived from C-moments in a mean squared error sense. Although it may be argued, from the limited experiments that we have performed that, when outliers are present in the data, C-moments derived from pdf's computed from L-moment constraints show somewhat less bias and a smaller variance than C-moments computed from samples, the difference is certainly not dramatic.

The fact that the L-moment derived pdf's are superior estimates of the true pdf's is interesting and may be of importance when the shape of the pdf is the object of the experiment. The Laplace pdf, for example, serves as a good illustration of the different estimates. It is clear that the L-moment derived pdf shows much more decisively than the conventional estimate that the pdf may be Laplacian. Such a conclusion can, of course, be then used to parametrically fit an actual Laplacian pdf to the data and thereby obtain estimates of the C-moments. This conclusion is also supported by the mixture of two distributions example, which demonstrates that L-moments are better for characterizing the main portion of the distribution than the C-moments.

An observation worth making that stems from our work, and is well demonstrated by our first three examples, is that the degree to which L-moments are superior estimates to C-moments, for a given small data sample, depends to a large extent on the pdf itself. i.e., the robustness of L-moments is, to some extent, data dependent.

In our limited experience, application of L-moment derived pdf's to field data appears to show strong advantages. The derived pdf's are certainly more consistent when various logs from the same area are compared. The same observation applies when short data lengths from the same log are considered in order to avoid possible non-stationarity. In the case of field data that included an imposed outlier, L-moment statistics were considerably more robust than the C-moment counterparts.

References

Domenico PA, Schwartz FW (1997) Physical and Chemical Hydrogeology. Wiley Interscience

Landwehr JM, Matalas NC, Wallis JR (1979) Probability-weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. Water Resour. Res. 15, 1055–1064

Gini C (1912) Vaariabilità e mutabilità, contributo allo studio delle distribuzioni e delle relazione statistiche. Studi Economico-Giurdici dell' Università di Cagliari 3, 1–158

Gouveia W, de Moraes FS, Scales JA (1998) Entropy, Inference and Information, submitted to J. Geophysical Int.

Hosking JRM (1990) L-moments: Analysis and estimation of distributions using linear combinations of order statistics. It J. R. Stat. Soc., Ser. B 52, 105–124

Jaynes ET (1957) Information theory and statistical mechanics. Physical Review 106, 620–630

Kaigh WD, Driscoll MF (1987) Numerical and data summary using O-statistics. American Statistician 296, 1702–1706

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical Recipes in FORTRAN: the Art of Scientific Computing. 140–146, Cambridge University Press

Rietsch E (1977) The maximum entropy approach to inverse problem. J. Geophysics 42, 489–506

Rosa ALR, Ulrych TJ (1991) Processing via spectral modeling. Geophysics 56, 1244–1251

Royston P (1992) Which measures of skewness and kurtosis are best? Statistics in Medicine 11, 333–343

Sacchi MD (1996) Aperture compensated Radon and Fourier transforms, Ph.D. thesis, UBC, Vancouver, Canada

Shannon CE (1948) A mathematical theory of communication. Bell Systems Technical J. 27, 379–423

Silverman BW (1986) Density Estimation for Statistics and Data Analysis, 46–47, Chapman & Hall, London

Shore JE, Johnson RW (1981) Properties of cross-entropy minimization. IEEE Transactions on Information Theory IT-27, 472–482

Sudicky EA (1986) A natural gradient experiment on solute transport in a sand aquifer: Spatial variability of hydraulic conductivity and its role in the dispersion process. Water Resour. Res. 24(7), 2069–2082

Ulrych TJ (1998) The "whiteness" hypothesis: reflectivity, inversion, chaos and Enders. submitted to Geophysics

Velis DR, Ulrych TJ (1996) Simulated annealing wavelet estimation via fourth-order cumulant matching. Geophysics 61, 1939–1948

Vogel RM (1995) Recent advances and themes in hydrology. Rev. Geophys. 33, Suppl.

Wang QJ (1996) Direct sample estimators of L moments. Water Resour. Res. 32, 3617–3619

Woodbury AD (1989) Bayesian updating revisited. Math. Geology 21(3), 285–308

Woodbury AD, Sudicky EA (1991) The geostatistical characteristics of the Borden Aquifer. Water Resour. Res. 27(4), 533–546

Woodbury AD, Ulrych TJ (1998) Minimum relative entropy and probabilistic inversion in groundwater hydrology. Stoch. Hydrol. & Hydraulics 12, 317–358