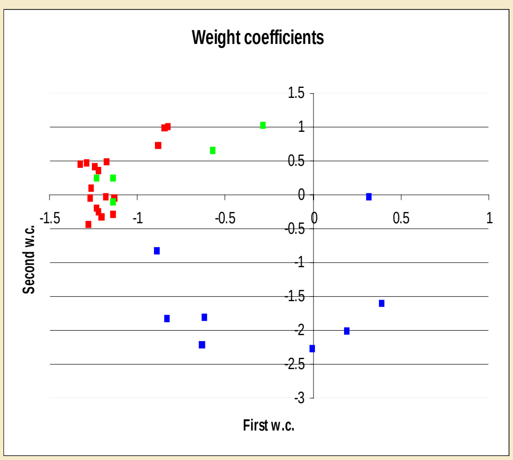
ARIDE

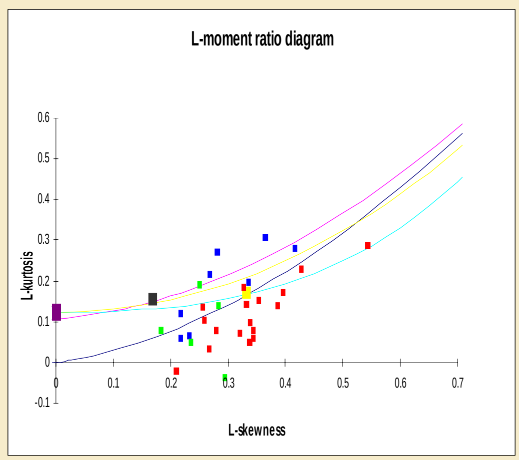
Assessment of the Regional Impact of Droughts in Europe

Technical Report No. 2

Methods for Regional Classification of Streamflow Drought Series:

The EOF Method and L-moments





October 1999

Department of Geophysics, University of Oslo
Norwegian Water Resources and Energy Directorate

NVE

Technical Report No. 2

Methods for Regional Classification of Streamflow Drought Series: The EOF Method and L-moments

Lena M. Tallaksen & Hege Hisdal

Technical Report to the ARIDE project No.2:

Supplement to Work Package 2 Hydro-meteorological Drought
Activity 2.3 Regional Drought Characteristics

© Department of Geophysics, University of Oslo, P.O. Box 1022 Blindern, N-0315 Oslo, Norway.

Contents

1. Objectives

In work package 2 (Hydrometeorological droughts) the statistical properties of long time series of precipitation and streamflow are analysed and emphasis is given to the regional aspect of hydrometeorological droughts. Two regionalisation tools are applied and compared, empirical orthogonal functions (EOFs) and L-moments. The regional characteristics of meteorological and hydrological droughts are compared, and homogeneous regions with respect to statistical properties (the EOF method) and frequency distribution (L-moments) of the extreme values of drought duration and deficit volume are identified. The method of EOFs is also, in combination with various interpolation techniques, applied as a tool to estimate the spatial distribution of drought characteristics, allowing the spatial extent of the drought to be mapped.

Section 2 and 3 of this technical note present the method of EOFs and L-moments, respectively. Both methodology details (theory and estimation procedures) as well as regionalisation approaches are discussed. In Section 4 application of the two methods are demonstrated using results from work package 2.3 (Regional drought characteristics), and a discussion of the suitability of the two methods for regional analysis is included.

2. The EOF Method

2.1 Introduction

The method of Empirical Orthogonal Functions (EOFs) also referred to as Principal Component Analysis has been described in many textbooks and articles. For a detailed description see Essenwanger (1976). There is a long tradition for applying the method in statistical analyses. The method is suitable for studies of variation patterns in time and space and for concentration of information in large data sets. Meteorologists started to use the EOF method for studies of variation patterns in temperature, precipitation and pressure fields (Grimmer, 1963; Stidd, 1967; Holmström & Stokes, 1978). The method has been applied by hydrologists from the late seventies, for example by Gottschalk & Krasovskaia (1979) for interpolation of water balance elements; by Bartlein (1982) for studies of anomaly patterns in monthly discharge in the USA and southern Canada; and later by Hisdal & Tveito (1992; 1993) for extension of runoff series and combined with kriging to generate runoff at ungauged locations. In Krasovskaia & Gottschalk (1995) the method was applied to study regional drought characteristics.

2.2 Method

The principle behind the EOF method applied in time series analysis is a linear transformation of spatially correlated series from a region into two sets of orthogonal and thus uncorrelated functions. The result is a set of series describing the temporal variations in the original data set and a set of series describing the spatial variability in the region from which the original series are collected. A short description of the method based on Hisdal & Tveito (1993) is given below. In this work the data series studied were daily streamflow records (1931-1980) from the southern part of Norway.

Let $Q(u_i,t)$ refer to measured streamflow at a station i among $N(i=1,\ldots,N)$ in a certain domain Ω. u corresponds to a two-dimensional plane $u=(x,y)$ and t is the time.

To suppress the influence of catchment area in the EOF analysis, a standardisation giving the runoff series zero mean and standard deviation equal to unity gives the best results (Haan, 1977):

$$X(u_i,t) = \frac{Q(u_i,t) - \overline{Q}_i}{s_{Q_i}}; \qquad i = 1,\ldots N \tag{1}$$

where:

$X(u_i,t)$ are standardised streamflow series

$\overline{Q}_i$ is the mean value of the series i

s_{Q_i} is the standard deviation of series i

The linear transformation can then be described as:

$$X(u,t) = \sum_{j=1}^{N} h_j(u)\beta_j(t) \tag{2}$$

where:

$h_j(u)$ are weight coefficients (elements of eigenvectors) describing the transformation

$\beta_j(t)$ are the EOFs describing the variation common to all original series

In a continuos formulation it can be shown that the eigenfunctions for the domain Ω are the solution to Fredholms homogeneous integral equation (Davenport & Root, 1958):

$$\int_{\Omega} R(u,u')h_j(u')du' = \mu_j h_j(u) \tag{3}$$

where:

$R(u,u')$ is the correlation function of the process X(u, t)

μ_j are the eigenvalues of the correlation matrix of the process X(u, t)

The EOFs can be obtained from an orthogonal projection of $X(u,t)$ on the eigenfunctions:

$$\beta_j(t) = \int_{\Omega} X(u,t)h_j(u)du \tag{4}$$

The following double orthogonality properties are then valid:

$$\int_{\Omega} h_j(u) h_k(u) du = \delta_{jk} \quad (5)$$

$$\mathrm{E}\left[\beta_{\mathrm{j}}(t)\beta_k(t)\right] = \mu_j \delta_{jk} \quad (6)$$

where:

δ_{jk} is the Kronecker delta

An estimation of the weight coefficients $h_j(u)$ requires a numerical solution. A simple way to approximate equation (3) is suggested by Holmström & Stokes (1978):

$$\sum_{l=1}^{N} R(u_i, u_l) h_j(u_l) = \mu_j h_j(u_i) \quad (7)$$

Equation (4) can be approximated by:

$$\beta_j(t) = \sum_{i=1}^{N} X(u_i, t) h_j(u_i) \quad (8)$$

Furthermore, the standardised river flow of a catchment can be described as:

$$X'(u_i, t) = \sum_{j=1}^{M} h_j(u_i) \beta_j(t) \quad (9)$$

A few of the EOFs ($\beta_j(t)$) will contain most of the variations in the original streamflow series. They are arranged in descending order according to the proportion of variance explained by each function. Often most of the variance in $X(u,t)$ can be described by a few EOFs (e.g., Hisdal & Tveito, 1993), and redundant information can be removed using only a few of the EOFs ($M<N$). The original streamflow records are totally described if all EOFs are applied ($M=N$).

The EOFs describe the temporal variation in streamflow in the region from which the original streamflow records were collected, and are common to all the initial series. In figure 1 an illustration of the three first EOFs based on daily streamflow records from a region with dominant snowmelt maximum flow in spring can be seen. The main cause of streamflow variability in the region, the snowmelt flood, can clearly be identified in the first EOF. The second EOF could be interpreted in terms of a delay in the timing of the spring flood, whereas the interpretation of the third EOF is more difficult. Interpretation of the physical meaning of EOFs where the original series have no main cause of variability is difficult (see figure 2, section 4).

The $h_j(u)$ describe the spatial variation of the original series, and can thereby be applied in a regionalisation study as described in section 2.3.

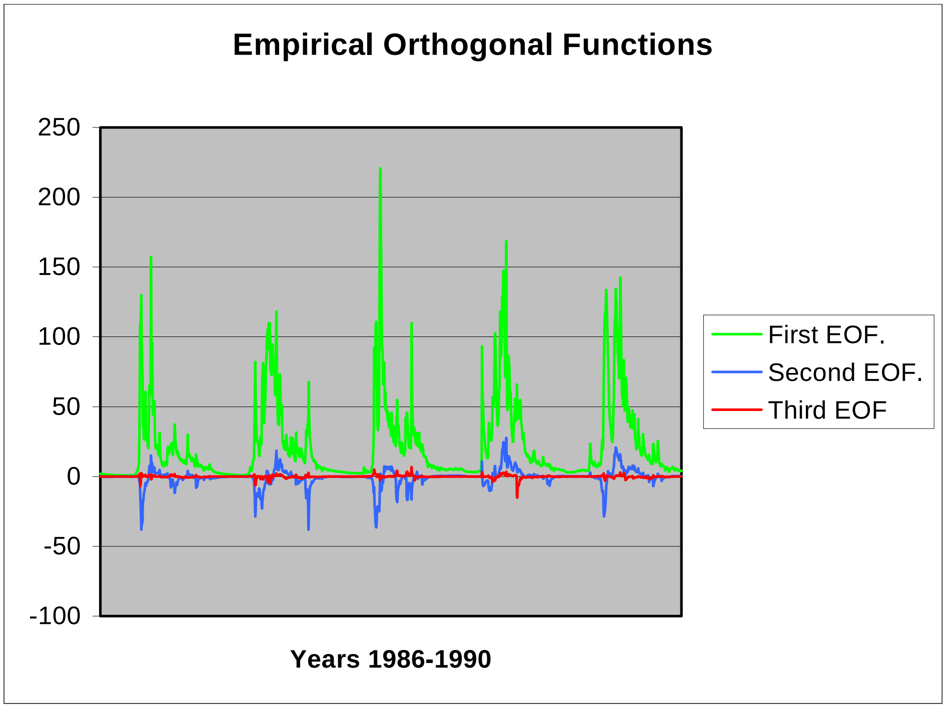


Fig. 1: The three first EOFs based on daily streamflow from a region with snowmelt flood.

The fact that these series $\beta_j(t)$ and the series $h_j(u)$ are referred to by different names by different authors might lead to confusion. In the following are some examples including the purpose for which the method was applied:

- Gottschalk (1985) applied the EOF method as a tool for regionalisation. In his article the temporal series, $\beta_j(t)$, are referred to as EOFs or amplitude functions and the spatial series, $h_j(u)$, as weight coefficients or elements of eigenvectors.
- Rao & Hsieh (1991) adopted the method for estimation of variables at ungauged locations. They regarded $h_j(u)$ as spatial EOFs and $\beta_j(t)$ as temporal EOFs.
- Lins (1985) studying interannual streamflow variability in the United States, used the term principal components for the $h_j(u)$ and component scores for the $\beta_j(t)$.
- Pandžić & Trninić(1991) presented a comprehensive consideration of the pattern of monthly streamflow, precipitation and water balance in a river basin in Yugoslavia. They applied the term principal component loadings for the $h_j(u)$ and principal component scores for the $\beta_j(t)$.
- Creutin & Obled (1982) analysing rainfall fields, named $h_j(u)$ eigenfunctions and $\beta_j(t)$ coefficients of the expansion.

2.3 Regionalisation

The weight coefficients, $h_j(u)$, describe the contribution of the EOFs to the original time series and can be regarded as the spatial component of the process $X(u,t)$. They vary between the initial series but are constant in time. The main common property of the initial series is reflected in the first EOF (ref. figure 1). The proportion of this main feature needed to obtain the original series at individual observation points, is reflected in the first weight coefficient. The second largest contribution to the total variability of the original series is described by the second EOF, and it's contribution to the original series is reflected through the second weight coefficient.

If a reasonable proportion of the total variability is described by the two first EOFs, a two-dimensional plot of $h_j(u)$ for $j=1$ and $j=2$ on a scatter diagram can serve to identify groups of stations with equal properties (e.g. figure 3, section 4). Other graphical presentations might allow for a regional classification based on more than two weight coefficients. The method is therefore applicable for regionalisation purposes. Gottschalk (1985) applied the method as a tool for regionalisation of daily streamflow in Sweden. The same approach was applied in a comparative study of different regionalisation methods for streamflow in Norway (Hisdal & Tveito, 1990).

The EOF method applied for regionalisation can be classified as a statistical method. As for cluster analysis the basis is a type of measure that reflects the distance between two points. In case of the EOF method the correlation coefficient is regarded as the distance function. As opposed to cluster analysis the group structure formed by the weight coefficients has to be assessed by eye. The groups identified in the scatter plot do not necessarily form geographically coherent regions on a map.

One advantage applying the EOF method is the fact that redundant information influencing the original series has been removed. The regionalisation is based only on differences in the main characteristics of the original series. This could give more distinct groups of stations. A disadvantage is, however, the complicated physical interpretation of the EOFs. The clusters obtained might therefore be difficult to explain. As with many other regionalisation procedures a limitation of the method is the requirement of the initial series to cover a common time period and be without missing values.

3 L-moments

3.1 Introduction

Hosking's (1990) method of L-moments has found widespread application in statistical analyses of hydrological data. L-moments are weighted linear sums of the expected order statistics and are analogous to conventional moments used to summarise the statistical properties of a probability function or an observed data set. The use of L-moments covers the characterisation of probability distributions, the summarisation of observed data samples, the fitting of probability distributions to data, the testing of hypotheses about distribution form and the identification of homogeneous regions. Gottschalk & Weingartner (1998) introduced an approach based on expected order statistics and L-moments for analysing the appearance of simple scaling and multiscaling in regional data sets on floods.

3.2 ***Method***

3.2.1 Definition

Let X be a real-valued random variable with cumulative distribution function $F(x)$ and quantile function $x(F)$, and let $X_{(1:n)} \leq X_{(2:n)} \leq \ldots \leq X_{(n:n)}$ be the order statistics of a random sample of size n drawn from the distribution of X. The first four L-moments are then defined as:

$$\begin{aligned}
\lambda_1 &= E[X_{(1:1)}] \\
\lambda_2 &= \tfrac{1}{2} E[X_{(2:2)} - X_{(1:2)}] \\
\lambda_3 &= \tfrac{1}{3} E[X_{(3:3)} - 2X_{(2:3)} + X_{(1:3)}] \\
\lambda_4 &= \tfrac{1}{4} E[X_{(4:4)} - 3X_{(3:4)} + 3X_{(2:4)} - X_{(1:4)}]
\end{aligned} \tag{10}$$

where E stands for expectation. The first moment equals the mean ($E[X]$), and the second moment is a measure of variation based on the expected difference between two randomly selected observations. It is common to standardize moments with higher order to make them independent of the unit of measurement of X. L-moment ratios, the L-coefficient of variation (τ_2), L-skewness (τ_3) and L-kurtosis (τ_4), are defined as:

$$\begin{aligned}
\tau_2 &= \frac{\lambda_2}{\lambda_1} \\
\tau_3 &= \frac{\lambda_3}{\lambda_2} \\
\tau_4 &= \frac{\lambda_4}{\lambda_2}
\end{aligned} \tag{11}$$

Skewness describes the relative asymmetry of a distribution, whereas kurtosis indicates the thickness of a distributions tail (peakedness). L-moment ratios are bounded, and the value of $\tau_r = \lambda_r/\lambda_2$ for $r \geq 3$ lies between –1 and +1.

Generally, the L-moments of X are defined to be the quantities:

$$\lambda_r \equiv r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E[X_{(r-k:r)}] \tag{12}$$

where r is the order of moment (r = 1,2,…), and

$$\binom{r-1}{k} = \frac{(r-1)!}{k!(r-1-k)!} \tag{13}$$

is the number of combinations of any k items from $(r-1)$ items.

The 'L' in L-moment emphasizes that λ_r is a linear function of the expected order statistics $E[X_{(r-k:r)}]$. The expectation of an order statistic is given as (e.g. Gibbons, 1985):

$$E\left[X_{(r:n)}\right] = \frac{1}{B(r, n-r+1)} \int_0^1 x(F) F^{r-1} [1-F]^{n-r} dF \tag{14}$$

From this the first L-moments are derived as (Hosking, 1990):

$$\begin{aligned}
\lambda_1 &= E[X] = \int_0^1 x(F) dF \\
\lambda_2 &= \tfrac{1}{2} E\left[X_{(2:2)} - X_{(1:2)}\right] = \int_0^1 x(F)(2F-1) dF \\
\lambda_3 &= \tfrac{1}{3} E\left[X_{(3:3)} - 2X_{(2:3)} + X_{(1:3)}\right] = \int_0^1 x(F)(6F^2 - 6F + 1) dF \\
\lambda_4 &= \tfrac{1}{4} E\left[X_{(4:4)} - 3X_{(3:4)} + 3X_{(2:4)} - X_{(1:4)}\right] = \int_0^1 x(F)(20F^3 - 30F^2 + 12F - 1) dF
\end{aligned} \tag{15}$$

Wang (1997) introduced LH moments, a generalization of L-moments that are based on linear combinations of higher-order statistics. The first moment is defined as the expectation of the largest value in a sample of size $\eta+1$, and provides a measure of the location of a distribution. Generally, the moment of order r considers the rth largest value in a subsample of size $r+\eta$. When $\eta = 0$ the LH moments become identical to ordinary L-moments. As η increases, LH moments (labeled Lη- moments) reflect more and more the upper part of distributions and larger events in the data. The method reduces the influence that small sample events may have on the estimation of large return period events, but is, however, sensitive to the determination of the appropriate value of η, the subsample size chosen. LL-moments can similar be defined considering the rth smallest value in a subsample of size $r+\eta$.

3.2.2 L-moments estimators

L-moments are defined for a probability distribution (ref. Eq. 15), but in practice estimation of L-moments must be made from a random sample drawn from an unknown distribution. Let $x_{(1:n)} \leq x_{(2:n)} \leq \ldots \leq x_{(n:n)}$ be the ordered sample from the observed sample of size n.

<u>U-statistics</u>

The L-moment of order r, λ_r, is a function of the expected order statistics of a sample of size r, and can be estimated using a U-statistic (Hoeffding, 1948). A U-statistic is the corresponding function of the sample order statistics averaged over all subsamples of size r which can be drawn from the observed sample of size n. For the first four sample L-moments we get:

$$
\begin{aligned}
l_1 &= n^{-1}\sum_i x_i \\
l_2 &= \frac{1}{2}\binom{n}{2}^{-1}\sum_{i>j}\sum\left(x_{(i:n)} - x_{(j:n)}\right) \\
l_3 &= \frac{1}{3}\binom{n}{3}^{-1}\sum\sum_{i>j>k}\sum\left(x_{(i:n)} - 2x_{(j:n)} + x_{(k:n)}\right) \\
l_4 &= \frac{1}{4}\binom{n}{4}^{-1}\sum\sum_{i>j>k>l}\sum\sum\left(x_{(i:n)} - 3x_{(j:n)} + 3x_{(k:n)} - x_{(l:n)}\right)
\end{aligned} \tag{16}
$$

<u>PWMs</u>

It is not necessary to iterate over all subsamples of size r, which can be quite large even for a relatively small sample size n. The L-moment statistic can be expressed explicitly as a linear combination of order statistics of a sample of size n (Hosking, 1990). The order statistics ($X_{(i:n)}$) are commonly expressed in terms of probability weighted moments (PWMs), and procedures based on PWMs and L-moments are equivalent. Greenwood et al. (1979) defined PWMs to be the quantities:

$$
M_{p,r,s} \equiv E\left\{X^p\left[F(X)\right]^r\left[1 - F(X)\right]^s\right\} \tag{17}
$$

where p, r and s are real numbers. They adopted the convention:

$$
M_{1,r,0} = \beta_r = E\left\{X\left[F(X)\right]^r\right\} \tag{18}
$$

for use in statistical inference procedures. A simple, but biased estimator of β_r can be obtained using a plotting position estimator of $F(x_{(i:n)})$, $p_{(i:n)}$ (Stedinger et al., 1993):

$$
b_r^* = n^{-1}\sum_{i=1}^{n} x_{(i:n)}\left[p_{(i:n)}\right]^r \tag{19}
$$

Estimators of L-moments obtained using unbiased PWM estimators, b_r, are given by:

$$
\begin{aligned}
&l_r = \sum_{k=0}^{r-1} p^*_{r-1,k}\, b_k \\
&\textit{where} \\
&b_r = n^{-1}\sum_{i=1}^{n}\frac{(i-1)(i-2)\ldots(i-r)}{(n-1)(n-2)\ldots(n-r)}\, x_{(i:n)} \\
&p^*_{r,k} = (-1)^{r-k}\binom{r}{k}\binom{r+k}{k}
\end{aligned} \tag{20}
$$

The first four L-moments are calculated as:

$$\begin{aligned}
\lambda_1 &= \beta_0 \\
\lambda_2 &= 2\beta_1 - \beta_0 \\
\lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0 \\
\lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0
\end{aligned} \tag{21}$$

Replacing sample estimators (b_r) into equations (21) provides the corresponding estimates for the L-moments. A library of FORTRAN subroutines useful for L-moment analyses is provided by Hosking (1991), and Stedinger et al. (1993) describe how to obtain the software by e-mail.

Direct estimators

Wang (1996) provides direct estimators of L-moments which eliminate the need for introducing PWMs. The estimation procedure follows closely the definition of L-moments by covering all possible combinations in a more efficient way. For the sample value $x_{(i:n)}$ there are *(i-1)* values $\leq x_{(i:n)}$ and *(n-i)* values $\geq x_{(i:n)}$, and for each subsample of size *r*, the number of values drawn from each of these categories are considered. The first four direct estimators are given by:

$$\begin{aligned}
l_1 &= \binom{n}{1}^{-1} \sum_{i=1}^{n} x_{(i:n)} \\
l_2 &= \frac{1}{2}\binom{n}{2}^{-1} \sum_{i=1}^{n} \left[\binom{i-1}{1} - \binom{n-i}{1} \right] x_{(i:n)} \\
l_3 &= \frac{1}{3}\binom{n}{3}^{-1} \sum_{i=1}^{n} \left[\binom{i-1}{2} - 2\binom{i-1}{1}\binom{n-i}{1} + \binom{n-i}{2} \right] x_{(i:n)} \\
l_4 &= \frac{1}{4}\binom{n}{4}^{-1} \sum_{i=1}^{n} \left[\binom{i-1}{3} - 3\binom{i-1}{2}\binom{n-i}{1} + 3\binom{i-1}{1}\binom{n-i}{2} - \binom{n-i}{3} \right] x_{(i:n)}
\end{aligned} \tag{22}$$

A Fortran subroutine is included which allows the readers to calculate the L-moments using the derived direct estimators.

Plotting-position estimators

A plotting position ($p_{(i:n)}$) is a distribution-free estimator of $F(x_{(i:n)})$. It provides a way to estimate quantities of the form $\int x(F)\eta(F)dF$, where $\eta(F)$ is a function of *F* alone. Equation (15) shows that λ_r is of this form, and it can be estimated by (Hosking, 1990):

$$l_r^* \equiv \sum_{i=1}^{n} \left(\sum_{k=0}^{r-1} p_{r-1,k}^* (p_{i:n})^k \right) x_{(i:n)} \tag{23}$$

which is equivalent to using the PWM estimator in (19) in combination with the L-moment estimator in (20). This approach does not provide unbiased estimators of λ_r. The choice $p_{(i:n)} = (i-0.35)/n$ has proved to give good results for the generalized Pareto, GEV and Wakeby distributions (Hosking, 1990).

<u>Comparison of estimators</u>
The numerical values of the unbiased sample estimators using U-statistics, PWMs and direct estimators are the same. (Landwehr *et al.*, 1979) recommend the use of biased estimates of PWMs and L-moments, since such estimators often produce quantile estimates with lower root-mean-square error than unbiased alternatives. Although there is no theoretical reason for preferring plotting-position estimators, experience has shown that they sometimes yield better results when a distribution is fitted to data at a single site. The unbiased estimators are recommended for calculating L moment diagrams and for use with regionalisation procedures where unbiasedness is important (Stedinger et al., 1993). Unbiased estimators are preferred because they have less bias for estimating L-moment ratios (Vogel & Fennessey, 1993).

3.2.3 Advantages of using L-moments
L-moments provide a unified approach to statistical inference for complete samples from continuos univariate distributions. The main advantage of L-moments over conventional product moments is that L-moments, being linear function of the data, suffer less from the effect of sample variability. For example, the second order product moment (standard deviation) and L-moment both measure the difference between two randomly drawn elements of a distribution. However, in case of standard deviation (σ^2) more weight is given to the largest differences as these are squared:

$$\lambda_2 = ½E[X_{(1:2)}-X_{(2:2)}] \qquad \sigma^2 = ½E[X_{(1:2)}-X_{(2:2)}]^2 \qquad (24)$$

The product moment-based measure of skewness is similar very sensitive to the extreme tails of the distribution (differences are now cubed), and is therefore difficult to estimate accurately when the distribution is markedly skew. Vogel & Fennessey (1993) demonstrated that sample product estimators of standard deviation and skewness exhibited significantly bias, even for extremely large sample sizes. The study by Sankarasubramanian & Srinivasan (1999), suggests, however, that product moments are preferable at lower skewness, particularly for smaller samples, while L-moments are preferable at higher skewness, for all sample sizes.

L-moment estimators, when compared with those of maximum likelihood estimators, usually show the method of L-moments to be reasonable efficient (Hosking, 1990). Hosking et al. (1985) reported that L-moment estimators had a lower root-mean-square error for the GEV distribution than the maximum likelihood estimates, and similar results were obtained by (Hosking & Wallis, 1987) for the generalized Pareto distribution.

Generally, L-moments are more robust to extreme values in the data and enable more secure inferences to be made from small samples about an underlying probability distribution (Hosking, 1990). A distribution might be specified by its L-moments even if some of its conventional product moments do not exist, and such a specification is always unique, which is not true for product moments (Hosking, 1990). Stedinger et al. (1993) provide formulas for the parameters of several distributions in terms of sample L-moments. The advantages offered in hypothesis testing, boundedness of moment ratios and identification of distributions have been discussed in detailed by Hosking (1986, 1990). The advantages of L-moment ratio estimators over product moment estimators are thoroughly discussed by Vogel & Fennessey (1993).

3.3 Regionalisation

In a regional analysis data sets from several sites are included and homogeneous regions in terms of the variable being studied are identified if possible. L-moment ratio diagrams permit the identification of homogeneous regions or grouping of catchments with respect to the distribution properties of the variable of interest. L-moment ratios (coefficient of variation, skewness and kurtosis) are analogous to ordinary product moment ratios plotted to yield an L-moment ratio diagram. The diagram summarises basic properties of theoretical probability distributions and observed samples, and has shown to be a powerful tool to depict contrasts between different samples (Hosking & Wallis, 1993; Vogel & Fennessey, 1993).

Figure 4 (section 4) shows an L-moment ratio diagram where L-skewness is plotted against L-kurtosis for observed drought samples and theoretical distributions. L-moments are here estimated using unbiased estimates of the PWMs. Two parameter distributions will show as points in the diagram, whereas three parameter distributions are depicted as curves. Ratio diagrams using L-coefficient of variation and L-skewness can similar be applied as a mean to obtain a general overview of the statistical properties of the sample observations (Gottschalk et al., 1997; Tallaksen & Hisdal, 1997).

L-moments are therefore well adapted to regional frequency analysis, a ‘region’ here meaning a group of catchments which is assumed to have data drawn from the same frequency distribution. Hosking & Wallis (1993) describe three statistics based on L-moments that are useful in regional frequency analysis: a discordancy measure, for identifying unusual sites in a region; a heterogeneity measure, for assessing whether a proposed region is homogeneous; and a goodness of fit measure for assessing whether a candidate distribution provides an adequate fit to the data. The suggested tests provide objective tools for the decisions involved in regional frequency analysis, and have been applied for homogeneity testing using annual maximum streamflow drought data in New-Zealand (Clausen & Pearson, 1995).

4 The EOF method and L-moments applied on drought data

The regional characteristics of severe seasonal droughts have been analysed by looking at the extreme value properties of annual maximum series (AMS) of drought characteristics using both the L-moment diagram and the EOF method (Tallaksen & Hisdal, 1997). A Nordic data set of 52 catchments covering 60 years of daily flow data was used to derive the AMS of drought duration and deficit volume using the threshold level approach. L-moment diagrams showed that the generalised Pareto distribution gave the best overall fit to the annual and summer drought samples. Deficit volume exhibited a more long-tailed behaviour than the distributions obtained for duration, and a lower variance cover was found for deficit volume using the EOF method. The two methods provided virtually the same conclusions with regard to clustering of catchments, and large scale trends were found which confirmed a regional pattern.

The low number of stations in the study hampered, however, a consistent regional comparison between catchments with the same summer season. In a new data set the number of stations was increased at the expense of the length of the observation series, and only stations with dominant summer low flow were included. This data set covers mainly the southern part of Scandinavia, i.e. Denmark and southern Sweden, but also catchments located along the coastline of Norway. The 30 series cover a common 30-year period (1965-1994). Annual maximum drought volumes for the summer season (May 1 – November 31) were derived from daily streamflow records applying the 70 percentile from the flow duration curve as a threshold level. Application of the EOF method and L-moments as tools for regionalisation are demonstrated below for this new data set.

Figure 2 presents an example of the three first EOFs. Together they cover about 68% of the total variability. The large negative value of the first EOF in 1976 can be interpreted as the severe drought this year that also covered large parts of the Nordic countries. It was followed by a general wet period in the 80's and subsequent dryer 90's. A physical interpretation of the second and third EOF is problematic.

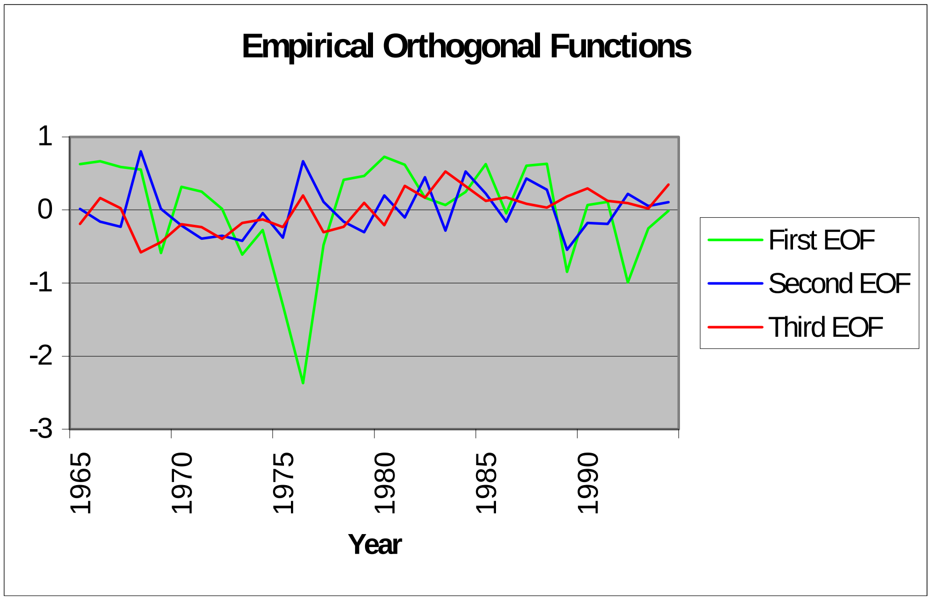


Fig. 2: The three first EOFs based on annual maximum deficit volume.

In figure 3 the two first weight coefficients corresponding to the first and second EOF in figure 2 are shown. Stations are labelled separately by country as a first step to judge whether or not it is possible to identify different geographical regions in the sample. A clear clustering of stations can be seen, and as illustrated in the figure, this can to a large extend be explained by geographical location.

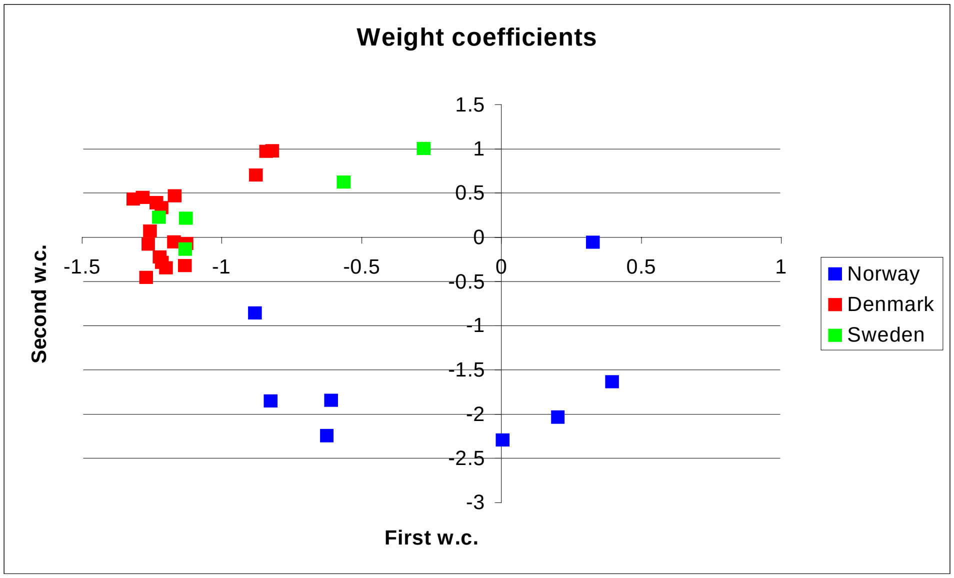


Fig. 3: First and second weight coefficient based on annual maximum deficit volume.

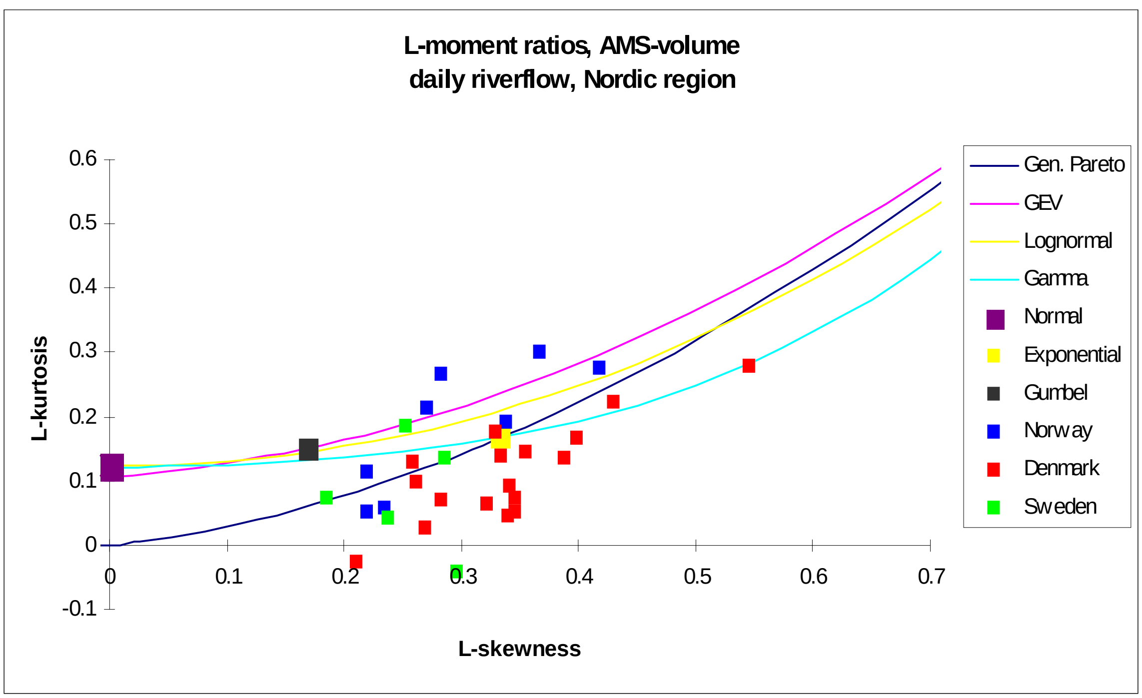


Fig. 4: L-moment diagram for annual maximum series of deficit volume.

Figure 4 shows the corresponding sample L-moments plotted together with several theoretical distribution functions. Again the generalised Pareto distribution seems to give the best overall fit to the data. The stations are, as in figure 2, labelled by country, and similar grouping of stations can be seen in the L-moment ratio diagram. The Norwegian stations are, however, clearer identified as a cluster in the EOF plot. In case of L-moments objective methods for delineating homogeneous regions are available as described in section 3.3. A more detailed presentation and discussing of the results will be given in the ARIDE annual report 1999.

References

Bartlein, P.J. (1982) Streamflow anomaly patterns in the USA and southern Canada 1951-70. *J. Hydrol.* **57** (1/2), 49-63.

Clausen, B. & Pearson, C.P. (1995) Regional frequency analysis of annual maximum streamflow drought. *J. Hydrol*., **173**, 111-130.

Creutin, J.D. & Obled, C. (1982) Objective analyses and mapping techniques for rainfall fields: an objective comparison. *Wat. Resour. Res.* **18** (2), 413-431.

Davenport, W.B. & Root, W.L. (1958) *An introduction to the theory of random signal and noise.* McGraw-Hill, New York, USA.

Essenwanger, O. (1976) *Applied statistics in atmospheric science.* Developments in Atmospheric Science **4A**, Elsevier, Amsterdam, the Netherlands.

Gibbons, J.D. (1985) Non parametric statistical inference. Marce Dekker, Inc., New York.

Gottschalk, L. (1985) Hydrological regionalization of Sweden. *Hydrol. Sci. J.* **30** (1), 65-83.

Gottschalk, L. & Krasovskaia, I. (1979) Synthetic approach to regional hydrology and physiography. SMHI, Research and Training Department, Norrköping, Sweden.

Gottschalk, L. & Weingartner, R. (1998) Distribution of peak flow derived from a distribution of rainfall volume and runoff coefficient, and a unit hydrograph. *J. Hydrol*., **208**, 148-162.

Gottschalk, L., Tallaksen, L.M. & Perzyna, G. (1997) Derivation of low flow distribution functions using recession curves. *J. Hydrol.* **194**, 239-262.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C. & Wallis, J.R. (1979) Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Wat. Resour. Res*., **15**, 1049-1054.

Grimmer, M. (1963) The space-filtering of monthly surface anomaly data in terms of pattern using empirical orthogonal functions. *Quart. J. Roy. Met. Soc.* **89**, 395-408.

Haan, C.T. (1977) *Statistical methods in hydrology.* The Iowa State University Press, Ames, Iowa, USA.

Hisdal, H. & Tveito, O.E. (1990) Regionalisation and the EOF-method (Regioninndeling med henblikk på EOF-metoden, in Norwegian), UiO-NVE, Oppdragsrapport 4-90, Oslo, Norway.

Hisdal, H. & Tveito, O.E. (1992) Generation of runoff series at ungauged locations using empirical orthogonal functions in combination with kriging. *Stochastic Hydrol. Hydraul.* **6,** 255-269.

Hisdal, H. & Tveito, O.E. (1993) Extension of runoff series using empirical orthogonal functions. *Hydrol. Sci. J.* **38** (1/2), 33-49.

Hoeffding, W. (1948) A class of statistics with asymptotically Normal distribution. *Ann. Math. Statist.*, **19**, 293-325.

Holmström, I. & Stokes, J. (1978) Statistical forecasting of sea level change in the Baltic. *SMHI Rapporter*, Nr RMK9, Norrköping, Sweden.

Hosking, J.R.M. (1986) The theory of probability weighted moments. *Research Report RC12210*, IBM Research, Yorktown Heights.
Hosking, J.R.M. (1990) L-moments: Analysis and estimation of distributions using linear combinations of order statistics. J. R. Statist. Soc., Ser. B, 52, 105-124.
Hosking, J.R.M. (1991) Fortran routines for use with the method of L-moments. *Research Report RC17097*, IBM Research, Yorktown Heights.
Hosking, J.R.M. & Wallis, J.R. (1987) Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, **29**, 339-349.
Hosking, J.R.M. & Wallis, J.R. (1993) Some statistics useful in regional frequency analysis. *Wat. Resour. Res*., **29(2)**, 271-281.
Hosking, J.R.M., Wallis, J.R. & Wood, E.F. (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, **27**, 251-261.
Krasovskaia, I & Gottschalk, L. (1995) Analysis of regional drought characteristics with empirical orthogonal functions. In: Z. W. Kundezewicz (Ed) New uncertainty concepts in hydrology and water resources, International hydrology series, Cambridge university press, 163-167.
Landwehr, J.M., Matalas, N.C. & Wallis, J.R. (1979) Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Wat. Resour. Res*., **15(5)**, 1055-1064.
Lins, H.F. (1985) Interannual streamflow variability in the United States based on Principal Components, *Wat. Resour. Res.* **21** (5), 691-701.
Pandžić, K. & Trninić, D. (1991) Principal component analysis of the annual regime of hydrological and meteorological fields in a river basin, *Int. J. Clim.* **11**, 909-922
Rao A.R. & Hsieh, C.H. (1991) Estimation of runoff at ungauged locations by empirical orthogonal functions, *J. Hydrol.* **123**, 51-67.
Sankarasubramanian, A. & Srinivasan, K. (1999) Investigation and comparison of sampling properties of L-moments and conventional moments. *J. Hydrol*., **218**, 13-34.
Stedinger, J.R., Vogel, R.M. & Foufoula-Georgiou, E. (1993) Frequency analysis of extreme events. In: D.R. Maidment (Editor), *Handbook of Hydrology*. McGraw-Hill, New York , p. 18.41.
Stidd, C.K. (1967) The use of eigenvectors for climate estimates. *J. Appl. Meteorol.* **6**, 255-264.
Tallaksen, L.M. & Hisdal, H. (1997) Regional analysis of extreme streamflow drought duration and deficit volume. Proceedings of the 3rd International conference on FRIEND, 1-4 Oct. 1997, Postojna, Slovenia, *IAHS Publ*., 246, 141-150.
Vogel, R.M. & Fennessey, N.M. (1993) L-moment diagrams should replace product moment diagrams. *Wat. Resour. Res*., **29(6)**, 1745-1752.
Wang, Q.J. (1996) Direct sample estimators of L moments. Water Res. Res., 32(12), 3617-3619.
Wang, Q.J. (1997) LH moments for statistical analysis of extreme events. Water Res. Res., 33(12), 2841-2848.