

# Mixed estimation methods for Halphen distributions with applications in extreme hydrologic events

**Fateh Chebana<sup>(1)\*</sup>**  
**Salaheddine El Adlouni<sup>(2)</sup>**  
**Bernard Bobée<sup>(1)</sup>**

(1) Institut National de la Recherche Scientifique, INRS-ETE Chaire en Hydrologie  
Statistique 490, rue de la couronne (Québec), Canada G1K 9A9

(2) Institut National de Statistique et d'Économie Appliquée, INSEA, Avenue allal El Fassi,  
Madinat AL Irfane, Rabat, 10100, Morocco

Institut National de la Recherche Scientifique, INRS-ETE  
Chaire en Hydrologie Statistique  
490, rue de la couronne (Québec), Canada G1K 9A9

**\*Corresponding author:** Tel: +1 (418) 654-2542  
Fax: +1 (418) 654-2600  
Email: [fateh\\_chebana@ete.inrs.ca](mailto:fateh_chebana@ete.inrs.ca)

Revised version submitted to: *Stochastic Environmental Research and Risk Assessment*.

May, 28<sup>th</sup> 2009

## **Abstract**

The Halphen family of distributions is a flexible and complete system to fit sets of observations independent and identically distributed. Recently, it is shown that this family of distributions represents a potential alternative to the generalized extreme value (GEV) distributions to model extreme hydrological events. The existence of jointly sufficient statistics for parameter estimation leads to optimality of the method of maximum likelihood (ML). Nevertheless, the ML method requires numerical approximations leading to less accurate values. However, estimators by the method of moments (MM) are explicit and their computation is fast. Even though MM method leads to good results, it is not optimal. In order to combine the advantages of the ML (optimality) and MM (efficiency and fast computations), two new mixed methods were proposed in this paper. One of the two methods is direct and the other is iterative, denoted respectively MMD and MMI. An overall comparison of the four estimation methods (MM, ML, MMD and MMI) was performed using Monte Carlo simulations regarding the three Halphen distributions. Generally, the MMI method can be considered for the three Halphen distributions since it is recommended for a majority of cases encountered in hydrology. The principal idea of the mixed methods MMD and MMI could be generalized for other distributions with complicated density functions.

## 1. Introduction

The main objective of frequency analysis of extreme events, and especially floods, is the assessment of the probability  $\Pr(X \geq x_T)$  of an event  $x_T$  to be exceeded. To this end, the first step is to fit a probability distribution  $F$  to a series of observed maximum annual flow, independent and identically distributed, and coming from a random variable  $X$ . Then, the event  $x_T$  corresponds to the quantile associated to a return period  $T$ .

Several distributions have been proposed to fit hydro-meteorological variables. Specific distributions are recommended in some countries, such as the Lognormal distribution (LN) in China (Bobée, 1999); the Generalized Extreme Value distribution (GEV) in the United Kingdom for local flood frequency analysis (NERC, 1975) and in the United States for precipitation (Willeke et al., 1995); and the Log-Pearson type 3 distribution (LP3) in the United States (US Water Resources Council, 1981). Halphen (1941) introduced the Halphen type A distribution (HA) to fit a large variety of data and to obtain additional flexibility compared to other distributions used at that time. Indeed, the HA distribution has a tail lighter than the Lognormal which fit a wide range of hydrological series (Morlat, 1951). According to Morlat (1956): "it is quickly observed that: the distributions of Type A showed to be adequate in many cases, but a significant number of sets of observations required new shapes, extending distributions of type A to represent, especially in the origin neighbourhood, different asymptotic behaviours." Thus Halphen has introduced the type B distribution (HB). To complete the Halphen system, the distribution of Halphen type inverse B (HIB) was proposed by Larcher as reported by Morlat (1956). The latter author also presented some statistical and mathematical properties of these distributions, which have been completed and detailed by Perreault et al. (1999a).

In order to estimate each Halphen distribution parameters, the ML method was proposed by Perreault et al. (1999b). The ML estimation is optimal for Halphen distributions since it is based on three statistics jointly sufficient. However, the resolution of the ML system is complex because it requires some manipulations of the derivatives of the normalization functions (Bessel and exponential factorial). To solve efficiently the ML system, a numerical method was presented by Perreault et al. (1999b). Note that the estimation by the ML method is based on numerical approximations which do not lead to exact values. Furthermore, the bias of the ML is asymptotically null but it may be large for small sample sizes which may lead to a large root-mean square-error. Consequently, despite the sufficiency of ML estimators, it is relevant to consider other estimation methods for small sample sizes encountered in hydrology. Fitzgerald (2000) proposed a method of moments (MM) to estimate the parameters of the HA distribution. An advantage of the method of moments is the explicit expressions of estimators with respect to the moments, thus facilitating its implementation and making computations faster. As a natural extension of the work of Fitzgerald (2000), Chebana et al. (2008a) developed a method of moments to estimate the parameters of HB and HIB distributions.

To obtain the ML estimators for each one of the three Halphen distributions  $f(x; m, \alpha, \nu)$ , Perreault et al. (1999b) carried out a partial maximization of the conditional log-likelihood function  $\log L(\nu | \alpha(\nu), m(\nu))$ . We propose two new mixed methods that combine the advantages of ML (optimality) and of MM (fast computations):

- The first method MMD is direct where the parameter  $\nu$  is estimated using MM equation which leads to a value of  $\nu$  close to the optimum value and then estimators of  $\alpha$  and  $m$  are obtained from the two corresponding equations of the ML system.
- The second method MMI is iterative. It consists in varying  $\nu$  from its estimator obtained by MM, and stop at the value of  $\nu$  which corresponds to maximizing the conditional log-

likelihood function  $\log L(v | \alpha(v), m(v))$ , and then  $\alpha$  and  $m$  are estimated from the two corresponding equations of ML.

On the other hand, note that the performances of the MM method for HA distribution have not been evaluated and have not been compared with ML. Chebana et al. (2008a) have made a comparison between the estimates by MM and ML for HB and HIB distributions. Thus in this work, we conduct a comprehensive comparison of the four proposed methods (MM, ML, MMD and MMI) for the three Halphen distributions (HA, HB and HIB).

Aside from the introduction, the paper is organized as follows. The main properties of Halphen distributions are presented in Section 2 as well as their place in the context of extreme hydrologic events. Section 3 concerns with the four estimation methods. A comparison, based on Monte Carlo simulations for the various estimation methods, is presented in Section 4. Results and a general discussion are developed in Sections 5 and 6, respectively. The last Section is dedicated to the study conclusions.

## **2. Halphen distributions: presentation and extreme hydrologic events**

### **background**

In the present section, we present briefly Halphen distributions and then we put them in the context of the extreme hydrologic events.

#### **2.1. Presentation**

Perreault et al. (1999a) presented a detailed review of mathematical and statistical properties of Halphen distributions: HA, HB and HIB which are summarized here.

### ***Halphen type A distribution (HA)***

The probability density function of the HA distribution is given by:

$$f_{HA}(x; m, \alpha, \nu) = \frac{1}{2m^\nu K_\nu(2\alpha)} x^{\nu-1} \exp\left[-\alpha\left(\frac{x}{m} + \frac{m}{x}\right)\right], \quad x > 0 \quad (1)$$

where  $m > 0$  is a scale parameter,  $\alpha > 0$  and  $\nu \in \mathbb{R}$  are shape parameters. The function  $K_\nu(\cdot)$  corresponds to the modified Bessel function of the second kind (Watson, 1996).

The non-centered moments of a random variable  $X$  following HA distribution can be determined by using the characteristic function. They exist for any real value of  $\nu$ ,  $\alpha > 0$  and any integer  $r$ , and are given by:

$$\left[\mu_r'\right]_{HA} = E[X^r] = m^r \frac{K_{\nu+r}(2\alpha)}{K_\nu(2\alpha)} \quad (2)$$

### ***Halphen type B distribution (HB)***

The probability density function of the HB distribution is given by:

$$f_{HB}(x; m, \alpha, \nu) = \frac{2}{m^{2\nu} ef_\nu(\alpha)} x^{2\nu-1} \exp\left[-\left(\frac{x}{m}\right)^2 + \alpha\left(\frac{x}{m}\right)\right], \quad x > 0 \quad (3)$$

where  $m > 0$  is a scale parameter,  $\alpha \in \mathbb{R}$  and  $\nu > 0$  are shape parameters. The normalization function  $ef_\nu(\cdot)$ , known as exponential factorial function, was defined by Halphen (1955) and is given by:

$$ef_\nu(\alpha) = 2 \int_0^\infty x^{2\nu-1} \exp(-x^2 + \alpha x) dx, \quad \nu > 0 \text{ and } \alpha \in \mathbb{R} \quad (4)$$

The  $r$ -order non-centered moments of HB distribution are given by:

$$\left[\mu_r'\right]_{HB} = E[X^r] = m^r \frac{ef_{\nu+r/2}(\alpha)}{ef_\nu(\alpha)} \quad (5)$$

Since the index  $\nu$  of the function  $ef_\nu(\cdot)$  must be positive, those moments exist for any real value of  $\alpha$  if  $(\nu + r/2) > 0$ . Hence, the non-centered moments of order  $r$  are defined for  $r$  greater than  $r_0 = \lceil -2\nu \rceil$  (where  $\lceil \cdot \rceil$  is the integer part).

### ***Halphen type Inverse B distribution (HIB)***

The probability density function of the HIB distribution is given by:

$$f_{HIB}(x; m, \alpha, \nu) = \frac{2}{m^{-2\nu} ef_\nu(\alpha)} x^{-2\nu-1} \exp\left[-\left(\frac{m}{x}\right)^2 + \alpha\left(\frac{m}{x}\right)\right], \quad x > 0 \quad (6)$$

where  $m > 0$  is a scale parameter,  $\alpha \in \mathbb{R}$  and  $\nu > 0$  are shape parameters.

The probability density of the HIB distribution (Equation 6) can be deduced from that of HB (Equation 3) by considering the following symmetry relationship:

*If a random variable  $X$  follows a  $f_{HB}(x; m, \alpha, \nu)$  distribution, then  $Y = 1/X$  follows  $f_{HIB}(y; m^{-1}, \alpha, \nu)$  distribution.*

Therefore, the non-centered moments of HIB can be deduced from equation (5) by replacing  $m$  by  $m^{-1}$  and  $r$  by  $-r$ , and are therefore given by:

$$\left[\mu_r'\right]_{HIB} = E[X^r] = m^r \frac{ef_{\nu-r/2}(\alpha)}{ef_\nu(\alpha)} \quad (7)$$

These moments exist for  $\nu - r/2 > 0$ , i.e., they are defined until the order  $r_0 = \lceil 2\nu \rceil$ .

Perreault et al. (1997) presented a detailed review of Halphen distributions and their limiting cases (Gamma and Inverse Gamma distributions), and studied, in detail, their mathematical and statistical properties. We can show (Perreault et al., 1997) that Halphen distributions belong to the exponential family of probability density functions with three

parameters and they admit a triplet of *jointly* sufficient statistics. A sufficient statistic is a function of observations which summarizes all the information related to the parameter to be estimated. Estimators based on sufficient statistics are efficient, i.e., they have a minimum variance regardless of sample size and are asymptotically unbiased. This is a very important property, especially in the cases of small samples encountered in hydrology. The Halphen distributions are the only ones with three parameters used in hydrology that have this enviable property. Distributions with three parameters are important, first because they are flexible and present several shapes of distributions and include several two and one parameter distributions, and on the other hand they obey to the parsimonious principle compared to the four and five parameters distributions (e.g. Kappa and Wakeby), see for instance Morlat (1956), page 26. In addition, three parameter distributions are important in tail distributions modeling as indicated in El Adlouni et al. (2008). Hence three parameters distributions represent good balance between flexibility and parsimony.

Adequate estimation of the parameters  $\underline{\theta}$  of a distribution  $F(.,\underline{\theta})$  is one of the most important problems of statistical inference, particularly in frequency analysis. Several three parameter distributions, used in hydrology, such as Log-Normal (LN), Log-Pearson type 3 (LP3), Pearson type 3 (P3), and Generalized Extreme Value (GEV) have a location parameter and the distribution support depends on this parameter. Hence, in this case, the estimation can not be optimal for a given finite sample size; and the maximum likelihood (ML) estimator properties are only asymptotic as has been shown by Bobée (1979) in the case of Pearson and Log-Pearson distributions. Indeed, the probability density function of such distributions can not be factorized in the form of a distribution belonging to the exponential class and it is not possible to find sufficient statistics. A distribution  $F(x;\theta_1,\theta_2,\theta_3)$  belongs to the exponential family if its probability density function can be expressed as:



$$f(x; \theta_1, \theta_2, \theta_3) = \exp \left\{ \sum_{i=1}^3 c_i(\theta_1, \theta_2, \theta_3) T_i(x) + d(\theta_1, \theta_2, \theta_3) + S(x) \right\} \quad (8)$$

where  $T_i(X), i = 1, \dots, 3$  represent the three jointly sufficient statistics. The Halphen distributions are of the form  $f(x; m, \alpha, \nu)$  and have three parameters: two shape parameters  $\alpha$  and  $\nu$  and a scale parameter  $m$  but no location parameter. We can then show (Perreault et al., 1997) that each Halphen distribution has three statistics jointly sufficient. Consequently, one can obtain optimal parameter estimation (i.e. minimum variance) for these distributions by the ML method regardless of the sample size.

The sufficient statistics for each one of the three Halphen distributions are related to the sample averages of different orders. Indeed, the associated sufficient statistics are  $H^{-1}$ ,  $\log G$  and  $A$  for HA;  $\log G$ ,  $A$  and  $Q$  for HB and  $IQ^{-1}$ ,  $H^{-1}$ , and  $\log G$  for HIB. The averages  $IQ^{-1}$ ,  $H^{-1}$ ,  $\log G$ ,  $A$  and  $Q$  correspond respectively to non-centered sample moments of order -2, -1, 0, 1 and 2. The quasi-zero moment ( $\underline{0}$ ) of a positive random variable  $X$  is given by  $\mu'_{\underline{0}} = \exp\{E[\log X]\}$ ; it represents the geometric mean (see Bobée and Ashkar, 1991, pp. 102 or Bobée et al., 1993, pp. 51). In the moment ratio diagram  $\delta_1 = \log\left(\frac{A}{G}\right)$  and  $\delta_2 = \log\left(\frac{G}{H}\right)$ , Morlat (1956) presented Halphen distributions as well as their limiting distributions (Harmonic, Gamma and Gamma Inverse) (Figure 1). This representation in the plan  $(\delta_1, \delta_2)$  shows that the Halphen family of distributions forms a complete system. Indeed, each point of the plan corresponds to one and only one theoretical Halphen distribution or one of their limiting cases. Other distributions, such as LP3, LN and generalized Gamma, can also be represented in the plane  $(\delta_1, \delta_2)$ , (Bobée et al., 1993). This diagram is similar to that of Pearson  $(\beta_1 = Cs^2, \beta_2 = Ck - 3)$  based on skewness and kurtosis coefficients  $Cs$  and  $Ck$  respectively. It is

shown that the Pearson family of distributions represents a complete system in the plane  $(\beta_1, \beta_2)$  (see Figure 1 in Bobée et al., 1993).

## 2.2. Halphen distributions in the context of extreme hydrologic events

Generally extreme hydrological events, such as floods, storms and droughts are associated with serious economic and social consequences. The prediction of such events is hence of high importance and requires the development of the appropriate models. Relating the magnitude of extreme events to their frequency of occurrence is the principal aim of frequency analysis. This relationship can be obtained through the use of probability distributions (see e.g., Chow et al., 1988). For a random variable  $X$  that represents the magnitude of an event that occurs at a given time and at a given site, the quantile function expresses the magnitude of the event in terms of its exceedence or non-exceedence probability. These probabilities are also associated to return periods. Hence, more precisely, the goal of frequency analysis is to obtain reliable estimates of the quantiles corresponding to return periods of specific relevance (Rao and Hamed, 2000). Explicitly, for floods, we are interested in the assessment of the probability  $\Pr(X \geq x_T)$  of an event  $x_T$  to be exceeded. Then, the event  $x_T$  corresponds to the quantile associated to a return period  $T$ , i.e.,  $x_T = F^{-1}(1 - 1/T; \underline{\theta})$  where  $\underline{\theta}$  is the vector of parameters. Since the interest is for large return periods (generally larger than 100 years),  $x_T$  is located in the tail of the distribution  $F$ . In flood frequency analysis, the notion of return period for hydrological extreme events is a common criterion. Physically, a return period is defined as the mean passed time between the occurrences of two events  $(X \geq x_T)$ , see e.g. Rao and Hamed (2000).

Several distributions have been proposed to fit hydrological variables as indicated in the introduction. Even though Halphen distributions are introduced in the 1940's, the literature

regarding Halphen distributions is very limited. Indeed, the use of Halphen distributions has been limited in practice because of numerical computational difficulties related to the normalization functions: modified Bessel function of the second kind (Abramowitz and Stegun, 1972) for HA and the exponential factorial function (Halphen, 1955) for HB and HIB. The papers Halphen (1941, 1955), Morlat (1956) and Perreault et al. (1999a,b) can be considered as the reference about the subject. Le Cam and Morlat (1949) applied HA and HB distributions on French rivers; Morlat (1951) discussed the asymptotic behaviour of HA and HB. Larivaille (1960) treated the estimation of HA and HB scale parameter with fixed shape parameters. Seshadri (1997) summarized Halphen distributions properties. It is important to mention that HA distribution corresponds to a reparameterization of the so-called Generalized Inverse Gaussian (GIG) distribution. The GIG distribution is studied by Sichel (1975), Barndorff-Neilsen and Halgreen (1977), Barndorff-Neilsen (1978), Blasield (1978) and Jorgensen (1982). In the context of GIG, Seshadri (1993) and Puig (2008) discussed briefly the HA distribution. In addition, Guillot (1964) proposed an extension of HA that includes the Galton-Gibrat distribution as a limiting case. On the other hand, the HB and HIB are not related to other known distributions as it is the case for HA. Morlat (1956) indicated that Halphen system is important similarly to Pearson system and are used especially to represent natural phenomena such as flows and precipitations.

The difficulties that prevented Halphen distributions to be used in practice are investigated by Perreault et al. (1999a,b) and the HYFRAN software (CHS, 2002) performs a complete frequency analysis for about twenty distributions often used in hydrology including the three Halphen distributions and their limiting cases.

The extreme value modeling is associated to the study of distributions generally with heavy tails. El Adlouni et al. (2008) presented a classification of the distributions commonly used in flood frequency analysis with respect to their tail behaviour. Note that quantile estimations associated to large return periods are significantly different for the considered distributions.

Several classes of tail distributions are identified and each distribution is assigned to a given class. It is shown that almost all distributions given in this classification to fit the annual maximum series belong to the class D of sub-exponential distributions or to the class C with regularly varying tail (Figure 2). In particular, the HA, HB distributions belong to the class D and the distribution HIB belong to the class C.

Recently, El Adlouni et al. (2009) presented a comparison between Halphen system and the widely used GEV system. This comparison involves historical, theoretical and practical aspects. El Adlouni et al. (2009) showed that Halphen system represents a potential alternative to the GEV system to model hydrological extreme series.

### **3. Detailed presentation of the estimation methods**

In this section the four methods of parameter estimation of the three Halphen distributions are presented. In the remainder, we denote by  $\theta_{XX}$  the estimator of a parameter  $\theta$  by a method XX.

#### **3.1. Method of moments (MM)**

The method of moments allows to obtain explicit expressions of the parameter estimators. We present in this section the expressions of the estimators for the three Halphen distributions. This approach corresponds to the framework of the generalized method of moments (Bobée and Ashkar, 1988).

##### **3.1.1 MM method for HA distribution**

For HA distribution, the MM estimators are proposed by Fitzgerald (2000). The obtained equations are based, on the one hand, on the differential equation satisfied by Bessel function and its recurring formula:

$$\alpha^2 \frac{\partial^2 K_\nu(\alpha)}{\partial \alpha^2} + \alpha \frac{\partial K_\nu(\alpha)}{\partial \alpha} - (\alpha^2 + \nu^2) = 0 \quad (9)$$

$$\frac{1}{\alpha} \frac{\partial}{\partial \alpha} (\alpha^\nu K_\nu(\alpha)) = -\alpha^{\nu-1} K_{\nu-1}(\alpha) \quad (10)$$

and on the other hand, on the expression of the moments for any order  $r$  given by equation (2).

For the three parameters, we obtain the following estimators:

$$m^2 = \frac{E(X^{-1}) \text{Var}(X) - E(X)(E(X)E(X^{-1}) - 1)}{E(X) \text{Var}(X^{-1}) - E(X^{-1})(E(X)E(X^{-1}) - 1)} \quad (11)$$

$$\alpha = \frac{m^{-1}E(X) - mE(X^{-1})}{m^{-2}\text{Var}(X) - m^2\text{Var}(X^{-1})} \quad (12)$$

$$\nu = \frac{(m^{-1}E(X) - mE(X^{-1}))^2}{m^{-2}\text{Var}(X) - m^2\text{Var}(X^{-1})} \quad (13)$$

In (11), (12) and (13), the theoretical moments are generally unknown. In order to obtain the estimators, those moments should be replaced by their estimates from the sample  $x_1, \dots, x_n$  of size

$n$ , ie.,  $E(X^r)$  is replaced by  $n^{-1} \sum_{i=1}^n x_i^r$ .

### 3.1.2 MM method for HB distribution

The MM method for HB and HIB distributions is proposed by Chebana et al. (2008a).

The normalization function  $ef_\nu(\cdot)$  satisfies the following recurrence relationship:

$$ef_{\nu+1}(\alpha) = \frac{\alpha}{2} ef_{\nu+1/2}(\alpha) + \nu ef_\nu(\alpha), \quad \nu > 0 \quad (14)$$

A more general form of equation (14) is given by:

$$ef_{\nu+r}(\alpha) = \frac{\alpha}{2} ef_{\nu+r-1/2}(\alpha) + (\nu+r-1) ef_{\nu+r-1}(\alpha), \quad (\nu+r) > 1 \quad (15)$$

This equation combined with expression (5) giving the non-centered moments of HB provides:

$$E\left(\frac{X^{2r}}{m^{2r}}\right) = \frac{\alpha}{2} E\left(\frac{X^{2r-1}}{m^{2r-1}}\right) + (\nu + r - 1) E\left(\frac{X^{2r-2}}{m^{2r-2}}\right), \quad (\nu + r) > 1 \quad (16)$$

Using this equation with  $r = 1/2; 1; 3/2$  leads to the following estimators:

$$\nu = \frac{1}{2} \frac{E(X)E(X^{-1}) \left[ E(X^3)E(X) - (E(X^2))^2 \right] - \text{Var}(X)(E(X))^2}{\left[ 1 - E(X)E(X^{-1}) \right] \left[ (E(X^2))^2 - E(X^3)E(X) \right] - (\text{Var}(X))^2} \quad (17)$$

$$m^2 = \frac{2\text{Var}(X)}{2\nu \left[ 1 - E(X)E(X^{-1}) \right] + E(X)E(X^{-1})} \quad (18)$$

$$\alpha = \frac{m \left[ 2\nu (E(X) - E(X^2)E(X^{-1})) + E(X^2)E(X^{-1}) \right]}{\text{Var}(X)} \quad (19)$$

As for HA distribution (see 3.1.1), in order to obtain the estimators  $\hat{m}$ ,  $\hat{\alpha}$  and  $\hat{\nu}$ , the theoretical moments should be replaced by their estimators in (17), (18) and (19).

### 3.1.3 MM method for HIB distribution

In order to obtain the MM estimators of the HIB distribution parameters, one can follow the same steps as those used for HB and consider equation (15) combined with (7) which gives the non-centered moments of HIB. We can also, and more simply, use the symmetry relationship between HB and HIB distributions (see Section 2.3) to obtain HIB estimators directly from those of HB. Thus, estimators of  $m$ ,  $\alpha$  and  $\nu$  can be obtained by substituting  $X^r$  by  $Y^{-r}$  and  $m$  by  $m^{-1}$  in (17), (18) and (19) which yield, respectively to:

$$\nu = \frac{1}{2} \frac{E(Y)E(Y^{-1}) \left[ E(Y^{-3})E(Y^{-1}) - (E(Y^{-2}))^2 \right] - \text{Var}(Y^{-1})(E(Y^{-1}))^2}{\left[ 1 - E(Y)E(Y^{-1}) \right] \left[ (E(Y^{-2}))^2 - E(Y^{-3})E(Y^{-1}) \right] - (\text{Var}(Y^{-1}))^2} \quad (20)$$

$$m^2 = \frac{2\nu \left[ 1 - E(Y)E(Y^{-1}) \right] + E(Y)E(Y^{-1})}{2\text{Var}(Y^{-1})} \quad (21)$$

$$\alpha = \frac{2\nu \left[ E(Y^{-1}) - E(Y^{-2})E(Y) \right] + E(Y^{-2})E(Y)}{m\text{Var}(Y^{-1})} \quad (22)$$

As for HA and HB distributions, theoretical moments should be replaced by their estimators in (20), (21) and (22) to obtain the estimators  $\hat{m}$ ,  $\hat{\alpha}$  and  $\hat{\nu}$ .

### 3.2. Maximum likelihood method (ML)

In the following, we present the essential elements of the ML method for each one of the three Halphen distributions. The complete presentation of the procedure is deferred in Appendix A. For more details, one can refer to Perreault et al. (1997). In short, the method consists, firstly, in estimating the parameters  $\alpha$  and  $m$  for fixed values of  $\nu$  in a given interval. Then, the partial log-likelihood function  $\log L(\nu | \alpha(\nu), m(\nu))$  is maximized to obtain the optimal estimator  $\hat{\nu}$  of  $\nu$ . Finally, the later estimator is substituted in the corresponding equations of the ML system to get the estimators  $\hat{\alpha}$  and  $\hat{m}$ . Note that this method has been implemented in the software HYFRAN (CHS, 2002). The properties of the ML estimators were studied by Chebana et al. (2006) using simulations to evaluate their performance on samples of finite sizes.

To solve the nonlinear system of equations (given below) and to determine the ML estimators, we proceeded in three stages (Perreault et al., 1997):

- a) The parameters  $\alpha$  and  $m$  are determined for fixed  $\nu$ , and are denoted by  $\alpha(\nu)$  et  $m(\nu)$ .
- b) The partial log-likelihood function  $\log L(\nu | \alpha(\nu), m(\nu))$  is maximized to determine the estimator  $\hat{\nu}$  of  $\nu$ .
- c) By substituting  $\hat{\nu}$  in  $\alpha(\nu)$  and  $m(\nu)$ , we obtain the estimators  $\hat{\alpha} = \alpha(\hat{\nu})$  and  $\hat{m} = m(\hat{\nu})$  of  $\alpha$  and  $m$ , respectively.

The steps (a) and (b) are developed in Appendix A for each Halphen distribution.

For the estimation of HA distribution parameters, the ML system is composed of the equations given by:

$$m \frac{K_{\nu+1}(2\alpha)}{K_{\nu}(2\alpha)} = \frac{1}{n} \sum_{i=1}^n x_i = A \quad (23)$$

$$\frac{1}{m} \frac{K_{\nu-1}(2\alpha)}{K_{\nu}(2\alpha)} = \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} = \frac{1}{H} \quad (24)$$

$$\log m + \frac{\partial K_{\nu}(2\alpha)/\partial \nu}{K_{\nu}(2\alpha)} = \frac{1}{n} \sum_{i=1}^n \log x_i = \log G \quad (25)$$

where  $A$ ,  $G$  and  $H$  are, respectively, the arithmetic mean, the geometric mean and the harmonic mean. This system reveals the three sufficient statistics related to the HA distribution.

The system of the ML estimator equations for the HB distribution is given by:

$$m \frac{ef_{\nu+\frac{1}{2}}(\alpha)}{ef_{\nu}(\alpha)} = \frac{1}{n} \sum_{i=1}^n x_i = A \quad (26)$$

$$m^2 \frac{ef_{\nu+1}(\alpha)}{ef_{\nu}(\alpha)} = \frac{1}{n} \sum_{i=1}^n x_i^2 = Q \quad (27)$$

$$\log m + \frac{\partial ef_{\nu}(\alpha)/\partial \nu}{2ef_{\nu}(\alpha)} = \frac{1}{n} \sum_{i=1}^n \log x_i = \log G \quad (28)$$

where  $A$ ,  $Q$  and  $G$  are, respectively, the arithmetic mean, the quadratic mean and the geometric mean. These moments represent the sufficient statistics related to the HB distribution .

As it was reported in Section 2.1, HIB distribution can be deduced directly from the HB one using the symmetry relationship. Thus, the system of equations for HIB may be obtained from the system of equations (26), (27) and (28) related to HB by substituting  $m$  by  $m^{-1}$  and  $x$  by  $x^{-1}$  . More precisely, the ML estimators are solutions of the following system:

$$\frac{1}{m} \frac{ef_{\nu+\frac{1}{2}}(\alpha)}{ef_{\nu}(\alpha)} = \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} = \frac{1}{H} \quad (29)$$



$$\frac{1}{m^2} \frac{ef_{\nu+1}(\alpha)}{ef_{\nu}(\alpha)} = \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i^2} = \frac{1}{IQ} \quad (30)$$

$$\log m - \frac{\partial ef_{\nu}(\alpha)/\partial \nu}{2ef_{\nu}(\alpha)} = \frac{1}{n} \sum_{i=1}^n \log x_i = \log G \quad (31)$$

where  $H$ ,  $IQ$  and  $G$  are, respectively, the harmonic mean, the inverse quadratic mean and the geometric mean. As for HA and HB distributions, this system reveals the sufficient statistics for HIB.

### 3.3. Direct mixed method (MMD)

This method is a combination of the two previous methods MM and ML. It consists in estimating  $\nu$  by MM method and then estimating  $\alpha$  and  $m$  using the corresponding equations of the ML system. Some required expressions related to the ML method are presented in Appendix A. For the case of HB and HIB distributions, in the MM method, the shape parameter  $\nu$  is estimated first before  $m$  and  $\alpha$ . Consequently, it can be directly substituted in the estimators of  $\alpha$  and  $m$  from ML. That is, for HB, to replace estimator (17) in equations (A.7) and (A.8), and for HIB, to replace estimator (20) in equations (A.13) and (A.14). However, for HA distribution, we obtain the explicit expression of the estimator of  $\nu$  by replacing that of  $m$ , expressed by (11) in equation (13). Indeed, the expression (13) can be re-written as:

$$\nu = \frac{\left(E(X) - m^2 E(X^{-1})\right)^2}{\text{Var}(X) - m^4 \text{Var}(X^{-1})} \quad (32)$$

By substituting the value of  $m^2$  given by formula (11) in both numerator and denominator of  $\nu$  (equation (32)), we obtain respectively:

$$\left(E(X) - m^2 E(X^{-1})\right)^2 = \left( \frac{\left(E(X)\right)^2 \text{Var}(X^{-1}) - \left(E(X^{-1})\right)^2 \text{Var}(X)}{E(X) \text{Var}(X^{-1}) - E(X^{-1}) \left[E(X)E(X^{-1}) - 1\right]} \right)^2$$

$$\text{Var}(X) - m^4 \text{Var}(X^{-1}) = \frac{\left[ (E(X))^2 \text{Var}(X^{-1}) - (E(X^{-1}))^2 \text{Var}(X) \right] \left[ \text{Var}(X) \text{Var}(X^{-1}) - [E(X)E(X^{-1}) - 1]^2 \right]}{\left[ E(X) \text{Var}(X^{-1}) - E(X^{-1}) [E(X)E(X^{-1}) - 1] \right]^2}$$

Finally, the explicit expression of  $\nu$  by MM method for HA is then given by:

$$\nu = \frac{(E(X))^2 \text{Var}(X^{-1}) - (E(X^{-1}))^2 \text{Var}(X)}{\text{Var}(X) \text{Var}(X^{-1}) - [E(X)E(X^{-1}) - 1]^2} \quad (33)$$

As mentioned in Section 3.1, the theoretical moments are to be replaced by their estimators in (33) in order to obtain the estimator  $\hat{\nu}$ . Thus, the estimators by the MMD method for the distribution HA can be obtained from (33) for  $\nu$ , and by replacing this in estimating equations (A.1) and (A.2), respectively,  $\alpha$  and  $m$  can be estimated.

This combined approach improves the resolution of the ML system, particularly in terms of computing time. Indeed, estimating  $\nu$  directly from the MM method avoids the partial maximization of the log-likelihood function with respect to  $\nu$  which may require a large number of iterations. The value of the estimator  $\nu_{MM}$  is close to the optimum and its value is explicitly given. Therefore, the MMD method allows to take into account the advantages of both methods, namely the speed of computations, the explicit formulation of MM estimators, and the sufficiency of the ML statistics to estimate  $\alpha$  and  $m$ . However, note that in MMD method, partial log-likelihood functions (equations (A.6), (A.12) and (A.16) for HA, HB and HIB respectively) are *not* explicitly used. In addition, the  $\nu_{MM}$  value does not exactly correspond to the optimum of  $\nu$ . The comparison in Section 4 is intended to determine whether this method leads to satisfactory results.

### 3.4. Iterative mixed method (MMI)

In order to accelerate numerical resolution of the maximum likelihood system, the proposed method reduces the variation interval that contains the optimal value of  $\nu$  that

maximizes  $\log L(\nu | \alpha, m)$ . Therefore, in a theoretical point of view, this method is as optimal as the ML method but less expensive in terms of computational time. It is another numerical way to resolve the same ML system.

In the ML method, the parameter  $\nu$  is varied within a predefined interval, namely  $]-U;U[$  for HA with  $U = AH^{-1}/[AH^{-1} - 1]$  (Appendix A.1);  $]0;V[$  for HB with  $V = 1/[2(Q/A^2 - 1)]$  (Appendix A.2) and  $]0;W[$  for HIB with  $W = 1/[2(H^2IQ^{-1} - 1)]$  (Appendix A.3). In the MMD method, the  $\nu_{MM}$  value obtained by MM is taken directly as the estimator of  $\nu$ , whereas in this proposed iterative method (MMI), the estimator  $\nu_{MM}$  is used as an initial value to the iterative process. Then, the variability direction of  $\nu$  is determined according to the sign of the variability of the partial log-likelihood function. Indeed, from the initial point, the values of  $\nu$  are increased by a fixed step (positive or negative) and we evaluate the partial log-likelihood function after having estimated  $\alpha$  and  $m$ . The change of sign of the derivative of the log-likelihood function, or the sign of the difference  $\Delta_k - \Delta_{k-1}$  (see algorithm below), indicates stopping of iterations and hence obtaining the maximum of the log-likelihood function. For example, Figure 3 illustrates for HA the variation range of  $\nu$  corresponding to this method and that of ML method.

More explicitly, the following algorithm shows how this method works:

1. Select a step  $s$ : for example take the same value used in ML, that is  $s = 0.1$ . This value is generally convenient for situations encountered in hydrology.
2. Initial stage:
  - a. Obtain  $\hat{\nu}_0 = \nu_{MM}$ : obtained by equations (33), (17) and (20) for HA, HB and HIB, respectively.
  - b. Compute  $\hat{\alpha}_0$  and  $\hat{m}_0$ : by substituting the corresponding  $\hat{\nu}_0$  in equations (A.1)-

(A.2) for HA, (A.7)–(A.8) for HB and (A.13)–(A.14) for HIB.

c. Compute the partial log-likelihood function  $L_0 = \log L(\hat{\nu}_0 | \hat{\alpha}_0, \hat{m}_0)$  using equations (A.6), (A.12) or (A.16) according to the target distribution.

3. Take  $\hat{\nu}_1 = \hat{\nu}_0 + s$  :

a. Compute  $\hat{\alpha}_1$  et  $\hat{m}_1$  that correspond to  $\hat{\nu}_1$  using the same equations as in stage (2.b).

b. Compute the partial log-likelihood function  $L_1 = \log L(\hat{\nu}_1 | \hat{\alpha}_1, \hat{m}_1)$  as in (2.c).

c. Compute the difference between  $L_1$  and  $L_0$ , denoted by  $\Delta_1 = L_1 - L_0$ .

4. Stage  $k$  ( $k = 1, 2, \dots$ ) :

a. If  $\Delta_1$  is positive :

While  $\Delta_k = L_k - L_{k-1}$  is positive, take  $\hat{\nu}_k = \hat{\nu}_{k-1} + s$  and the corresponding  $\hat{\alpha}_k$  and  $\hat{m}_k$  to evaluate  $L_k = \log L(\hat{\nu}_k | \hat{\alpha}_k, \hat{m}_k)$ .

b. If  $\Delta_1$  is negative :

While  $\Delta_k = L_k - L_{k-1}$  is negative, then take  $\hat{\nu}_k = \hat{\nu}_{k-1} - s$  and the corresponding  $\hat{\alpha}_k$  and  $\hat{m}_k$  to assess  $L_k = \log L(\hat{\nu}_k | \hat{\alpha}_k, \hat{m}_k)$ .

5. Stop iterations when the sign of  $\Delta_k$  changes. That means the maximum is reached.

6. Finally, the obtained estimators are  $\hat{\nu}_{MMI} = \hat{\nu}_{k^*-1}$ ,  $\hat{\alpha}_{MMI} = \hat{\alpha}_{k^*-1}$  and  $\hat{m}_{MMI} = \hat{m}_{k^*-1}$  where  $k^*$  is associated to the final step, i.e. it corresponds to the sign change of the derivative of the partial log-likelihood function.

Since the variation range of  $\nu$  is reduced in this method, it is possible to come as close as we want to the optimum by taking smaller value of the step  $s$ . Note that for comparison purposes, the value of  $s$  is taken to be the same for both ML and MMI.

When implementing the algorithm of the MMI method for each Halphen distributions, there are three elements that need to be clarified before proceeding to the comparison of the above methods:

- Although the estimator  $\nu_{MM}$  belongs to the variation interval of  $\nu$ , it is not necessarily a part of points of the grid of values employed in ML. It is likely that  $\nu_{MM}$  can be found between two points of this grid.
- In the MMI algorithm, the value of  $\nu$  of the last iteration is not at the maximum of the partial log-likelihood function. It corresponds to the change of sign variations of that function. Hence, the value maximizing the partial log-likelihood function is that associated with the iteration before the last one.
- One should ensure, at a given iteration  $k$ , that the value of  $m$  corresponding to the value of  $\alpha$  is obtained at the same iteration  $k$  which in turn corresponds to the value of  $\nu$  obtained at the iteration  $k$ .
- The iteration number is finite (and small), since the partial log-likelihood of Halphen distributions are strictly concave (Perreault et al., 1999b).

These points are illustrated with examples in Appendix C of Chebana et al. (2008b). In addition, some short simulations in Chebana et al. (2008b) illustrate the variation interval of the estimators of  $\nu$  using ML and MMI on the basis of one case of each distribution. It is clearly shown that the iteration number required for MMI is significantly smaller than the one required for ML. Indeed, on the basis of the simulations in Chebana et al. (2008b), the ML and MMI

methods require, respectively, 4 and 172 iterations for HA, 2 and 29 iterations for HB and 10 and 98 iterations for HIB.

#### 4. Comparison of the estimation methods using simulations

A comparison of the estimation methods is useful to evaluate the errors of estimation with respect to sample size and return period. The comparison is carried out using Monte Carlo simulations for the three Halphen distributions. For each one of the three distributions, we compare the four estimation methods MM, ML, MMD and MMI.

The selection of the studied cases is based on the values of the coefficients  $C_v$  and  $C_s$ . These coefficients can be obtained from the moment formulae (Perreault et al. 1999a). To determine the study area on the diagram ( $C_v$ ,  $C_s$ ), we considered series of annual maximum flow of 30 stations on rivers throughout the world. These series were drawn from the UNESCO website ([http://www.unesco.org/water/ihp/db/index\\_fr.shtml](http://www.unesco.org/water/ihp/db/index_fr.shtml)) and are relatively long with observation periods greater than 70 years. Statistical tests to check the iid property (independent and identically distributed) of this series were made in El Adlouni et al. (2007, 2008). In addition, for a sample size  $n$ , the coefficient  $C_s$  can be estimated by:

$$\hat{C}_s = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}} \quad (34)$$

It is well-known that  $\hat{C}_s$  is a biased estimator (Kirby, 1974). A classical correction is given by (Bobée and Ashkar, 1991, page 68):

$$\tilde{C}_s = \hat{C}_s \frac{\sqrt{n(n-1)}}{n-2} \left( 1 + \frac{8.5}{n} \right) \quad (35)$$

Figure 4 illustrates on the diagram ( $C_v$ ,  $C_s$ ) the location of the studied cases, the observed series with and without correction for  $C_s$ . One can see that the bias of  $C_s$  has no impact on the selection

of the studied cases, because the sample size of the observed series is large. Indeed, for  $n \geq 70$ , the correction coefficient in equation (35) is smaller than 1.15. In Figure 4, the obtained points for each observed series with and without bias correction are on a straight vertical line (same value of  $C_v$ ). In addition, Figure 5 presents on the  $(\delta_1, \delta_2)$  diagram the observed series as well as the studied cases. The selected cases corresponding to HIB are closer to the observed series in Figure 5 than in Figure 4, however, HB cases can be better justified using Figure 4. Figures 4 and 5 show that it is useful to select cases on the basis of both  $(C_v, C_s)$  and  $(\delta_1, \delta_2)$  diagrams. However, this issue is not common and deserves to be examined in further details in a future investigation.

The  $(C_v, C_s)$  values of the considered cases along with the corresponding parameter values for each distribution are indicated in Tables 1, 2 and 3 where four cases for each distribution are considered. Note that the scale parameter  $m$  is arbitrarily fixed at  $m = 100$ , since the coefficients  $C_v$  and  $C_s$  do not depend on  $m$ .

The simulation procedure that follows is similar to that described by Chebana et al. (2008a,b). Indeed, it is based on the generation of  $N = 1000$  samples of size  $n = 50$  and  $100$  using the algorithms proposed by El Adlouni and Bobée (2007). These algorithms are based on the acceptance-rejection method.

The performances of the estimators are evaluated using the relative bias (RB) and the relative root mean square errors (RRMSE). For a parameter  $\theta$  estimated by estimators  $\hat{\theta}_1, \dots, \hat{\theta}_N$ , the criteria RB and RRMSE are given respectively by:

$$RB = \frac{1}{N} \sum_{i=1}^N \left( \frac{\hat{\theta}_i - \theta_0}{\theta_0} \right) \quad (36)$$

$$RRMSE = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left( \frac{\hat{\theta}_i - \theta_0}{\theta_0} \right)^2} \quad (37)$$

## 5. Results

Since the main purpose of flood frequency analysis is the estimation of quantiles, we present results related to quantiles. However, results related to parameter estimation are presented and discussed in Chebana et al. (2008b). In this study, for each one of the three distributions HA, HB and HIB, we focus on the  $Q_T$  quantiles for return periods  $T = 10, 100$  and  $200$  years. The simulation results are presented in Tables 1, 2 and 3, respectively, for HA, HB and HIB distributions. It is important to notice that magnitudes of  $C_v$  and  $C_s$  are associated with the corresponding Halphen distribution. That is, from Figure 4, a low  $C_s$  value for HIB is large for HB (e.g., compare HIB1 and HB4).

In the following, we present results for each Halphen distribution (5.1, 5.2 and 5.3). In addition, general considerations are deferred in section 6.

### 5.1 HA distribution quantiles

The results for HA estimation quantiles are presented in Table 1. For the return period  $T = 10$  years, all methods lead to very similar results in all studied cases, especially for the RRMSE. Note that, as expected, the values of RB and RRMSE decrease with respect to  $n$ . In addition, the RB of  $Q_{10}$  varies between  $-2.93\%$  and  $3.25\%$ , while its RRMSE ranges from  $3.71\%$  to  $12.11\%$  and increases with respect to  $C_s$  for fixed  $n$ . One can see that, except in case 4 which corresponds to the highest values of the coefficients  $C_v$  and  $C_s$ , the quantile  $Q_{10}$  is overestimated by all methods.

When  $T = 100$  years, the results of the four methods are also very similar. It can be seen that the values of RB and RRMSE decrease with respect to  $n$ . The RB varies between  $-5.30\%$  and  $6.57\%$  with no monotonous behaviour with respect to  $C_v$  or  $C_s$ . However, the RRMSE varies



between 7.44% and 19.78% and it is increasing with respect to  $C_s$ . Moreover, the four methods overestimate  $Q_{100}$  except in the last case corresponding to the largest value of  $C_s$  (case 4).

For  $T = 200$  years, there is still similarity between the results of the four methods especially between ML and MMI. The RB and the RRMSE vary respectively between -5.34% and 7.88%, and between 12.07% and 22.50%. Unlike the RB behaviour, the RRMSE increases with respect to  $C_s$ . When  $n$  increases from 50 to 100, a decreasing of the RRMSE and a slight decreasing of RB can be observed. Moreover, as for  $T = 100$  years, the quantile  $Q_{200}$  is overestimated by all methods except in case 4.

In light of these results, one can conclude that, the four methods are very similar in estimating HA distribution quantiles. Consequently, the MMI method can be recommended because it combines theoretical optimality, accurate numerical results and computation speed. However, the MM method can also be recommended since it is the faster, the simplest and it leads to very close results to those of MMI.

## **5.2 HB distribution quantiles**

The estimation results of HB distribution quantiles are presented in Table 2. One can observe, in each case, that the sign of the RB is the same by all methods. Indeed, the RB is negative (underestimation) in the last three cases and it is positive (overestimation) only in case 1. Then, the four estimation methods underestimate HB quantiles except in case 1. For the considered three return periods, the results of the four methods are very similar and identical in some cases particularly if  $n = 100$  and  $T = 10$  or 100. Moreover, the values of the criteria resulting from ML and MMI are very similar; however, some differences are significant for large values of  $T$  and  $C_s$  (cases 3 and 4). Generally, one can observe that the BR (in absolute value) and the RRMSE increase with respect to  $C_s$  and they slightly decrease with respect to  $n$ .

For  $T = 10$  years, RB and RRMSE vary respectively between -3.09% and 0.57% and between 2.93% and 8.97%. For  $T = 100$  years, similarity between the methods decreases compared to those obtained for  $T = 10$  years. However, the results are comparable especially for the RRMSE. The RB varies between -8.85% and 1.12% and the RRMSE between 3.62% and 12.25%. When  $T = 200$  years, the RB and RRMSE of all methods have the same order of magnitude. In addition, in the last two cases, differences between results become more important, especially for the RRMSE. In these two cases, both MM and MMD methods lead to the lowest RB but accompanied with the biggest RRMSE. This means that the bias is not the principal source of variation. The variation in RB is between the limits -9.87% and 1.33% while for the RRMSE, the corresponding limits are 3.80% and 29.45%.

From these results, the MMD method can be recommended, except for high skewness cases to estimate quantiles of large return periods where MMI can be recommended instead.

### **5.3 HIB distribution quantiles**

Results related to HIB quantile estimation are presented in Table 3. For  $T = 10$  years, the results of the four methods are very similar especially for the RRMSE except for MMI in case 1. This exception is due to the parameter estimation for this case where it is shown in Chebana et al. (2008b) that the MMI estimation errors are the highest for the three parameters compared to the three other methods. The RB of  $Q_{10}$  estimation is very low and ranges between 0.08% and 1.42% and there is no monotonous trend with respect to Cs or Cv. The RRMSE varies between 2.64% and 6.79% with a slight increase with respect to Cs.

When  $T = 100$ , the results of the ML method are close to those of MMI, except in case 1. Compared to  $Q_{10}$ , the differences become larger between the results of the group MM and MMD and the group ML and MMI. An increasing trend of the bias is observed only for MM and MMD. The RB varies between the limits -0.01% and 6%. Apart some exceptions, for each method, the

RRMSE is increasing with respect to Cs, and its global variation is in the range from 4.58% to 17.60%.

The  $Q_{200}$  estimation results are very similar to those of  $Q_{100}$  in terms of behaviour. In the last three cases, ML and MMI results are similar and are better than those of MM and MMD. In the first case, the ML method leads to the lowest values of RB and RRMSE. The RB varies between -0.17% and 9.57% and the RRMSE varies between 5.23% and 25.84%.

For all considered return periods, the corresponding quantiles are generally overestimated by the four methods. According to the above results, one can conclude that for small return periods, the four estimation methods have similar performances except MMI for low skewness where it has less performance. For medium and large return periods, in cases of low skewness (in the HIB considered range), except MMI, the three other methods have approximately similar performances with slight superiority of ML. Hence, by taking into account other considerations (simplicity, optimality, rapidity), the MMD method is recommended for medium and large return periods, in cases of low skewness (in the HIB range) while in the other cases the MMI method can be recommended by taking into account results and calculation time.

In Tables 1, 2 and 3, maximum and minimum values of the RB and RRMSE are indicated for each Halphen distribution and each return period corresponding to quantile estimation by the four methods. From these values, no general conclusion can be drawn since all the methods have approximately similar performances.

## **6. General discussion**

In this section we present some common elements to all distributions as well as a summary of the obtained results. Generally, errors decrease when the sample size  $n$  increases for parameter estimation as well as for quantile estimation for a fixed return period. In addition, regarding quantile estimation, for fixed sample size, the criteria values increase with respect to

the return period  $T$ . For fixed return period and sample size, absolute values of the RB and the RRMSE are generally increasing with respect to  $C_s$  values (from case 1 to case 4), with some exceptions such as the RB in estimations related to HA. For each case of a given Halphen distribution, the gap between the results of the estimated quantiles using the four methods increases with respect to the return period but generally it remains low. Moreover, for each distribution, the results obtained from the four estimation methods are homogeneous in terms of behaviour.

In order to have a more clear understanding of the MMI estimation method, Figure 6 is produced by a simulation of 100 samples from case 3 of HA with sample size  $n = 100$ . The number of simulated samples here is for illustrative purpose and it is different from the one used in the described simulation study. The variation interval of the estimators of  $\nu$  using the different methods is shown in Figure 6. One can observe that the interval  $]-U;U[$  corresponding to ML method is very large for each sample and all estimators of  $\nu$  are almost merged. As a consequence, the interval where  $\nu$  varies is very tight for the MMI method since the estimator  $\nu_{MM}$  is very close to the optimum. This may explain the significant difference between the ML and MMI methods in terms of the average number of iterations necessary to achieve the optimum value. In this example, these averages are 4 and 172, respectively, for MMI and ML. Similar illustrations are presented in Chebana et al. (2008b) for both HB and HIB and conducting to the same conclusion.

In all cases, the errors expressed in terms of the criteria RB and RRMSE of quantiles are low in comparison with those parameters (see also Chebana et al., 2008b). This remark is not exclusive to Halphen family of distributions; it is generally observed in frequency analysis with distributions with three or more parameters. This effect can be explained by the compensation of parameter errors. Indeed, Based on the results of parameter estimation presented in Chebana et al.

(2008b), for a given case and a given sample size  $n$ , two important elements appear: (i) there is at least one parameter whose RB has different sign from the two others, and (ii) the RRMSE of the three parameters do not have the same order of magnitude. These two factors explain the compensation effect of parameter estimation errors affecting quantile estimation.

When considering computation simplicity and run speed of programs, the four methods are decreasingly rated as follows: MM, MMD, MMI and ML. In terms of theoretical optimality, based on the number of used sufficient statistics, they can be ordered as follows: ML and MMI, MMD and MM. With similar performances, this hierarchy is useful to select an estimation method.

The results of the comparative study can be summarized as follows:

- For HA distribution quantiles: despite the similarities between the four methods, the MMI method can be recommended because it combines theoretical optimality, numerical accuracy and fast computation. Nevertheless, in practical terms and with the same performance, MM method can also be recommended, thanks to its simplicity.
- For HB distribution quantiles: the MMD method can be recommended except in cases of large skewness (in the HB range) to estimate quantiles associated to large return periods. In such cases, MMI method may be recommended.
- For HIB distribution quantiles: for small return periods, the four methods have similar performances except MMI for very low skewness (in the HIB range). For medium and large return periods, in the case of very low skewness (in the HIB range), the MMD method is recommended while in other cases the MMI method can be recommended, taking into account results, its relative simplicity and required computation time.

Figure 7 summarizes and illustrates the recommended methods to estimate quantiles for each one of Halphen distributions. It concerns particularly estimations corresponding to  $T = 100$  and 200 years since these return periods are mostly used in practice. Estimations related to  $T = 10$  years are very similar by all compared methods. It is important to notice that the indicated boundaries in Figure 7 are based on the considered cases in the present study. However, it would be of interest to conduct a thorough study to define these boundaries in more efficient and general manner.

## 7. Conclusions

The present paper covers quantile estimation of the three Halphen distributions for which two new mixed estimation methods are proposed. From a theoretical point of view, the ML method is optimal. Nevertheless, resolving the corresponding non-linear system requires numerical approximations. Moreover, the maximization of the partial log-likelihood function requires a significant number of iterations. Furthermore, the bias of the ML is asymptotically null but it may be large for small sample sizes. Hence, it is relevant to consider other estimation methods for small sample sizes frequently encountered in several fields such as hydrology. In addition, the method of moments (MM) gives estimators with explicit expressions. Despite that it leads to good results, it is not optimal from a theoretical point of view since it is not based on the sufficient statistics of Halphen distributions. These reasons motivated the development of two new mixed methods, direct MMD and iterative MMI. The mixed methods combine the advantages of the methods MM (explicit expressions and rapidity) and ML (sufficiency).

A performance comparison of the four estimation methods (MM, ML, MMD and MMI) for each one of the three Halphen distributions was conducted using Monte Carlo simulations. The comparison showed homogeneity of the behaviour of all compared methods. The results of this comparison in terms of quantile estimation show for the HA distribution, great similarity of

the four methods; however, the MMI method is the most adequate in both theoretical point of view and result accuracy. Nevertheless, in practice and with the same performance, MM method may be recommended as well and MMD and ML should be avoided. For HB distribution, the MMD method can be recommended except in cases of large skewness (in the range of HB) to estimate quantiles of large return periods in which cases the MMI method may be recommended. For HIB distribution, the MMI method can be recommended considering results accuracy and computation time, except in the case of a very low skewness (in the range of HIB) where the MMD method can be recommended.

Finally, as a short conclusion, the MMI method can be considered for a majority of cases encountered in hydrology. In addition, it is worth to mention that the main idea of the mixed methods MMD and MMI can be conceptually applied to other distributions with several parameters and complicated forms such as distributions from the exponential family.

**NOTE:** Codes related to the presented estimation methods are available upon request from the authors.

### **Acknowledgment:**

The authors wish to thank the Natural Sciences and Engineering Research Council (NSERC) for financial support. The authors wish to thank the Editor-in-Chief and the two anonymous reviewers for their useful comments which led to the improvement of the paper.

## Bibliography

- Abramowitz, M. and Stegun, I. A. (1972) *Handbook of mathematical functions*. Dover, New York. 1006 pp.
- Barndorff-Nielsen, O. and Halgreen, C., (1977) Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 38(4): 309-311.
- Barndorff-Nielsen, O.E. (1978) Hyperbolic distributions and distributions on hyperbolae. *Scandinavian J. Statistics*, 5, 151-157.
- Blasiold, P. (1978) The shape of the generalized inverse Gaussian and hyperbolic distributions. Research Report No. 37. Dept. Theor. Statist., Aarhus University..
- Bobée, B. (1979) Comment on: The log-Pearson type 3 distribution: The T-year event and its asymptotic standard error by maximum likelihood theory. *Water Resources Research*, **15**, 189-190.
- Bobée, B. (1999) Extreme flood events valuation using frequency analysis: a critical review. *Houille Blanche*, **54** (7-8), 100-105.
- Bobée, B. and Ashkar, F. (1988) Generalized Method of Moments Applied to LP3 Distribution. *Journal of Hydraulic Engineering (ASCE)*, **114**, 899-909.
- Bobée, B. and Ashkar, F. (1991) *The Gamma Family and Derived Distributions Applied in Hydrology*. Littleton, Colorado: Water Resources Publications. 203 pp.
- Bobée, B.; Ashkar, F. and Perreault, L. (1993) Two kinds of moment ratio diagrams and their applications in hydrology. *Stochastic Hydrology and Hydraulics*, **7**, 41-65.
- Chebana, F.; El Adlouni, S. et Bobée, B. (2006) Propriétés des estimateurs du maximum de vraisemblance des paramètres et quantiles des lois de Halphen. [in French] *Research report*, I 219. INRS-ETE, Québec, Canada.



- Chebana, F.; El Adlouni, S. and Bobée, B. (2008a) Method of moments of the Halphen distribution parameters. *Stochastic Environmental Research and Risk Assessment*, **22**, 749-757, doi 10.1007/s00477-007-0184-4.
- Chebana, F.; El Adlouni, S. and Bobée, B. (2008b) Lois de Halphen : méthodes d'estimation mixtes et comparaison. [in French] *Research report*, R-994. INRS-ETE, Québec, Canada. 72 pp.
- Chow, V.T.; Maidment, D.R. and Mays, L.R. (1988) *Applied Hydrology*. McGraw-Hill, New York.
- CHS, Chaire en Hydrologie Statistique (2002) HYFRAN: Software for frequency analysis in hydrology. [in French] *Technical report*. INRS-Eau, Quebec, Canada.
- El Adlouni, S. and Bobée, B. (2007) Sampling techniques for Halphen distributions. *Journal of Hydrologic Engineering*, **12** (6) 592-604.
- El Adlouni, S.; Bobée, B. and Ouarda, T. B.M.J. (2008) On the tails of extreme event distributions in Hydrology. *Journal of Hydrology*, **355**, 16-33.
- El Adlouni, S. ; Chebana, F. and Bobée, B. (2009) Generalized Extreme Value vs. Halphen System: An explanatory study. *Journal of Hydrologic Engineering ASCE*. *Accepted*.
- Fitzgerald, D. L. (2000) Statistical aspects of Tricomi's function and modified Bessel functions of the second kind. *Stochastic Environmental Research and Risk Assessment*, **14**, 139-158.
- Guillot, P. (1964) Une extension des lois A de Halphen comprenant comme cas limite la loi de Galton-Gibrat. [in French] *Revue de Statistique Appliquée*, **12**, 63-73
- Halphen, E. (1941) Sur un nouveau type de courbe de fréquence. *Comptes Rendus de l'Académie des Sciences*, **213**, 633-635. Published under the name of « Dugué » due to war constraints.
- Halphen, E. (1955) Les fonctions factorielles. *Publications de l'Institut de Statistique de l'Université de Paris*, Vol. **IV**, Fascicule I. 21-39.

- Jørgensen, B. (1982) *Statistical properties of the generalized inverse Gaussian distribution*. Lecture Notes in Statistics, Springer, New York.
- Kirby, W. (1974) Algebraic boundedness of sample statistics. *Water Resour. Res.*, **10**(2), 220–222.
- Larivaille, P. (1960) Lois de Halphen: Estimation du paramètre d'échelle [in French]. *Publication Interne de L'E.D.F., Direction de L'Équipement*, 3 (194). EDF, Grenoble, France
- Le Cam, L. and Morlat, G. (1949) Les lois des débits des rivières françaises [in French]. *La Houille Blanche*. 1-7. No spécial B
- Morlat, G. (1951) Note sur l'estimation des débits de crues [in French]. *La Houille Blanche*, No spécial B, pp. 663-681.
- Morlat, G. (1956) Les lois de probabilité de Halphen [in French]. *Revue de Statistique Appliquée*, **3**, 21-43.
- NERC, Natural Environment Research Council (1975) *Flood Studies Report*, vol. 1, London.
- Perreault, L. ; Bobée, B. et Rasmussen, P.F. (1997) Les lois de Halphen. [in French] *Research report*, R-498, INRS-Eau, Université du Québec.
- Perreault, L.; Bobée. B. and Rasmussen. P.F. (1999a) Halphen Distribution System. I: Mathematical and Statistical Properties. *Journal of Hydrologic Engineering*, **4**, 189- 199.
- Perreault, L.; Bobée, B. and Rasmussen, P.F. (1999b) Halphen Distribution System. II: Parameter and quantile Estimation. *Journal of Hydrologic Engineering*, **4**, 200- 208.
- Puig, P. (2008) A note on the harmonic law: A two-parameter family of distributions for ratios. *Statistics and Probability Letters* 78 (3), 320-326.
- Rao, A.R. and Hamed, K.H. (2000) *Flood Frequency Analysis*. CRC Press, Boca Raton.
- Seshadri, V. (1993) *The Inverse Gaussian Distribution*. Clarendon Press, Oxford, U.K

- Seshadri, V. (1997) Halphen's laws. *Encyclopedia of Statistical Sciences*, 1, pp. 302-306. .S. Kotz and N. L. Johnson, eds.
- Sichel, H.S. (1975) On a distribution law for word frequencies *J. Am. Statistician Assn.*, 70. 542-547.
- U.S. Water Resources Council (1981) Guidelines for Determining Flood Flow Frequency. *Bull. 17B*, Washington DC.
- Watson, G.N. (1996) *A Treatise on the Theory of Bessel Functions*. 2<sup>nd</sup> Edition, Cambridge University Press. 812 pp.
- Willeke, G.E., Hosking, J.R.M., Wallis, J.R. and Guttman, N.B. (1995) The national drought atlas (draft), *IWR Rep. 94-NDS-4*, U.S. Army Corps of Eng., Fort Belvoir, Va.

## Appendix A: ML estimation

The object of this appendix is to provide some elements related to the ML estimation. These elements are required to the development of the MMD and MMI methods. Hereafter, we present in more details the three steps (a) and (b) given in Section 3.2. The main steps of the ML method, based on the previous numerical resolution approach for the three Halphen distributions, are illustrated in Figure A.1.

### A.1 ML for HA

a) For a fixed  $\nu$ ,  $\alpha(\nu)$  and  $m(\nu)$  are solutions of the system of equations (23) and (24)

which can be rewritten as:

$$D_A(\alpha(\nu), \nu) = \frac{A}{H} \quad (\text{A.1})$$

$$m(\nu) = \frac{A}{R_A(\alpha(\nu), \nu)} \quad (\text{A.2})$$

where

$$D_A(\alpha, \nu) = \frac{K_{\nu+1}(2\alpha)K_{\nu-1}(2\alpha)}{[K_\nu(2\alpha)]^2} \quad (\text{A.3})$$

and

$$R_A(\alpha, \nu) = \frac{K_{\nu+1}(2\alpha)}{K_\nu(2\alpha)} \quad (\text{A.4})$$

b) In order to determine the variation interval of  $\nu$ , we find signs of  $l'_{HA}(U)$  and  $l'_{HA}(-U)$

where  $l'_{HA}(\nu)$  is the derivative of the partial log-likelihood function  $\log L_{HA}$ , we have :

$$l'_{HA}(-U) = n \left[ \log \left( \frac{G}{H} \frac{1}{U} \right) + \Psi(U) \right] \quad \text{and} \quad l'_{HA}(U) = n \left[ \log \left( \frac{G}{A} U \right) - \Psi(U) \right] \quad (\text{A.5})$$

where  $\Psi(z) = \partial[\log \Gamma(z)] / \partial z$  is the digamma function and  $U = AH^{-1} / [AH^{-1} - 1]$ . Note that

$U > 1$ , since  $H \leq A$ . Hence:

- If  $l'_{HA}(U) > 0$  and  $l'_{HA}(-U) > 0$ , then the value of  $\nu$  that maximizes  $\log L_{HA}(\nu | \alpha, m)$  is greater than  $U$ , and in that case, the sample would better presented by a Gamma

distribution;

- If  $l'_{HA}(U) < 0$  and  $l'_{HA}(-U) < 0$ , then the value of  $\nu$  that maximizes  $\log L_{HA}$  is less than  $-U$ , and in that case, the sample would be better presented by an Inverse Gamma distribution.
- If  $l'_{HA}(U) < 0$  and  $l'_{HA}(-U) > 0$ , then the value of  $\nu$  that maximizes  $\log L_{HA}$  is in the interval  $]-U; U[$ ; and the ML estimators are solutions of the system composed by equations (A.1) and (A.2). In that case, the partial log-likelihood function is given by :

$$\log L_{HA}(\nu | \alpha, m) = n \left\{ \log \left[ \frac{G^{\nu-1}}{2m^\nu K_\nu(2\alpha)} \right] - \frac{\alpha [K_{\nu+1}(2\alpha) + K_{\nu-1}(2\alpha)]}{K_\nu(2\alpha)} \right\} \quad (\text{A.6})$$

## A.2 ML for HB

a) For a fixed  $\nu$ ,  $\alpha(\nu)$  and  $m(\nu)$  are solutions of the following system obtained from equations (26) and (27):

$$D_B(\alpha(\nu), \nu) = \frac{Q}{A^2} \quad (\text{A.7})$$

$$m(\nu) = \frac{A}{R_B(\alpha(\nu), \nu)} \quad (\text{A.8})$$

where

$$D_B(\alpha, \nu) = \frac{ef_{\nu+1}(\alpha)ef_\nu(\alpha)}{[ef_{\nu+1/2}(\alpha)]^2} \quad (\text{A.9})$$

and

$$R_B(\alpha, \nu) = \frac{ef_{\nu+1/2}(\alpha)}{ef_\nu(\alpha)} \quad (\text{A.10})$$

b) In order to establish the variation interval of  $\nu$ , we determine the sign of  $l''_{HB}(\nu)$ , the derivative of the partial log-likelihood function  $\log L_{HB}$ , evaluated at  $V = 1 / [2(Q/A^2 - 1)]$ , we obtain then:

$$l''_{HB}(V) = 2n \left[ \log \left( \frac{G}{A} 2V \right) - \Psi(2V) \right] \quad (\text{A.11})$$

Hence:

- If  $l'_{HB}(V) \geq 0$ , then the argument of the maximum of the partial log-likelihood function  $\log L_{HB}$  with respect to  $\nu$  is greater than  $V$ . In that case, the Gamma distribution is more appropriate to represent the sample;
- If  $l'_{HB}(V) < 0$ , then the maximum of the partial log-likelihood function  $\log L_{HB}$  with respect to  $\nu$  is reached in the interval  $]0;V[$ . In that case, the partial log-likelihood function is given by:

$$\log L_{HB}(\nu | \alpha, m) = n \left\{ \log \left[ \frac{2G^{2\nu-1}}{m^{2\nu} ef_{\nu}(\alpha)} \right] - \frac{ef_{\nu+1}(\alpha)}{ef_{\nu}(\alpha)} + \hat{\alpha} \frac{ef_{\nu+1/2}(\alpha)}{ef_{\nu}(\alpha)} \right\} \quad (\text{A.12})$$

### A.3 ML for HIB

**a)** For a fixed  $\nu$ ,  $\alpha(\nu)$  and  $m(\nu)$  are solutions of the following system obtained from equations (29) and (30):

$$D_B(\alpha(\nu), \nu) = \frac{H^2}{IQ} \quad (\text{A.13})$$

$$m(\nu) = HR_B(\alpha(\nu), \nu) \quad (\text{A.14})$$

where the functions  $D_B$  and  $R_B$  are given respectively by expressions (A.9) and (A.10).

**b)** To establish the interval in which belongs  $\nu$ , we determine the sign of  $l'_{HB}(W)$  which is the derivative of the partial log-likelihood function  $\log L_{HB}$  evaluated at

$W = 1 / [2(H^2 IQ^{-1} - 1)]$ , given by:

$$l'_{HB}(W) = 2n \left[ \log \left( \frac{H}{G} 2W \right) - \Psi(2W) \right] \quad (\text{A.15})$$

Therefore:

- If  $l'_{HIB}(W) \geq 0$ , then the value of  $\nu$  that maximizes the log-likelihood function  $\log L_{HIB}$  is greater than  $W$ . In this case, the Inverse Gamma distribution is more appropriate to represent the sample;
- If  $l'_{HIB}(W) < 0$ , then the maximum of function  $\log L_{HIB}$  with respect to  $\nu$  is reached in the interval  $]0;W[$ . In this case, the log-likelihood function is given by:

$$\log L_{HIB}(\nu | \alpha, m) = n \left\{ \log \left[ \frac{2m^{2\nu}}{G^{2\nu+1} ef_{\nu}(\alpha)} \right] - \frac{ef_{\nu+1}(\alpha)}{ef_{\nu}(\alpha)} + \hat{\alpha} \frac{ef_{\nu+1/2}(\alpha)}{ef_{\nu}(\alpha)} \right\} \quad (\text{A.16})$$

## List of Tables and Figures:

Tab. 1: RB and RRMSE of quantile estimation of HA distribution using the four methods (%)

Tab. 2: RB and RRMSE of quantile estimation of HB distribution using the four methods (%)

Tab. 3: RB and RRMSE of quantile estimation of HIB distribution using the four methods (%)

Fig. 1: Halphen distributions in the  $(\delta_1, \delta_2)$  diagram with  $\delta_1 = \ln\left(\frac{A}{G}\right)$  and  $\delta_2 = \ln\left(\frac{G}{H}\right)$  (From Morlat, 1956)

Fig. 2: Tail behaviour classification of distributions commonly used in hydrology (From El Adlouni et al., 2008).

Fig. 3: Illustration of interval variations of  $\nu$  using ML and MMI for HA distribution

Fig.4: Halphen distributions in the Cv-Cs diagram with the studied cases and the UNESCO series

Fig.5: The studied cases and the UNESCO series in the  $(\delta_1, \delta_2)$  diagram with  $\delta_1 = \log\left(\frac{A}{G}\right)$  and  $\delta_2 = \log\left(\frac{G}{H}\right)$

Fig. 6: Estimation of  $\nu$  and the corresponding empirical interval of variation for HA (case 3 and  $n = 100$ )

Fig. 7: Summary of conclusions related to quantile estimation results

Fig. A.1: ML estimation algorithm for parameters of Halphen distributions



**Tab. 1: RB and RRMSE of quantile estimation of HA distribution using the four methods (%)**

	$n = 50$								$n = 100$							
	RB				RRMSE				RB				RRMSE			
	MM	ML	MMD	MMI	MM	ML	MMD	MMI	MM	ML	MMD	MMI	MM	ML	MMD	MMI
<b>case 1</b>	$C_s = 0.88; C_v = 0.26$															
$T =$	$m = 100; \alpha = 7.00; \nu = -6.00$															
10	2.20	2.22	2.20	2.21	5.44	5.45	5.44	5.45	0.63	0.64	0.64	0.64	<u><b>3.71</b></u>	3.71	3.71	3.71
100	4.61	4.54	4.62	4.56	10.45	10.32	10.46	10.36	2.07	1.99	2.08	2.02	7.55	<u><b>7.44</b></u>	7.56	7.48
200	5.46	5.33	5.47	5.37	12.28	12.07	12.29	12.13	2.69	2.57	2.70	2.61	8.99	<u><b>8.82</b></u>	8.99	8.87
<b>case 2</b>	$C_s = 1.14; C_v = 0.51$															
$T =$	$m = 100; \alpha = 1.22; \nu = 2.82$															
10	1.07	1.26	1.21	1.26	8.14	8.16	8.15	8.16	1.06	1.15	1.13	1.15	6.00	6.01	6.01	6.01
100	6.12	6.34	6.38	6.33	14.35	14.28	14.57	14.29	3.84	4.01	3.95	4.02	9.91	9.95	9.99	9.94
200	7.58	7.78	<u><b>7.88</b></u>	7.78	16.70	16.51	17.00	16.51	4.59	4.79	4.72	4.79	11.19	11.23	11.30	11.22
<b>case 3</b>	$C_s = 1.59; C_v = 0.43$															
$T =$	$m = 100; \alpha = 2.40; \nu = -3.90$															
10	3.10	3.25	3.20	<u><b>3.25</b></u>	9.23	9.33	9.31	9.33	1.25	1.37	1.33	1.37	6.02	6.07	6.06	6.07
100	6.35	6.24	<u><b>6.57</b></u>	6.24	16.24	16.13	16.44	16.12	5.45	5.15	5.63	5.15	11.52	11.36	11.65	11.36
200	7.57	7.33	7.83	7.32	18.66	18.43	18.89	18.42	7.09	6.59	7.31	6.59	13.61	13.30	13.78	13.30
<b>case 4</b>	$C_s = 2.31; C_v = 0.71$															
$T =$	$m = 100; \alpha = 1.00; \nu = -1.00$															
10	<u><b>-2.93</b></u>	-2.26	-2.37	-2.26	<u><b>12.11</b></u>	12.10	12.10	12.10	-2.72	-2.15	-2.22	-2.15	8.58	8.57	8.57	8.57
100	<u><b>-5.30</b></u>	-4.61	-4.19	-4.62	19.43	19.48	<u><b>19.78</b></u>	19.48	-4.37	-3.79	-3.39	-3.80	14.21	14.11	14.54	14.12
200	<u><b>-5.34</b></u>	-4.75	-4.07	-4.76	21.91	21.90	<u><b>22.50</b></u>	21.91	-4.37	-3.88	-3.27	-3.88	16.22	16.01	16.72	16.01

Numbers underlined and in bold characters represent the minimum value of the criterion (RB or RRMSE) for given  $T$ . Numbers in bold and italic characters represent the maximum value of the criterion (RB or RRMSE) for given  $T$ .

**Tab. 2: RB and RRMSE of quantile estimation of HB distribution using the four methods (%)**

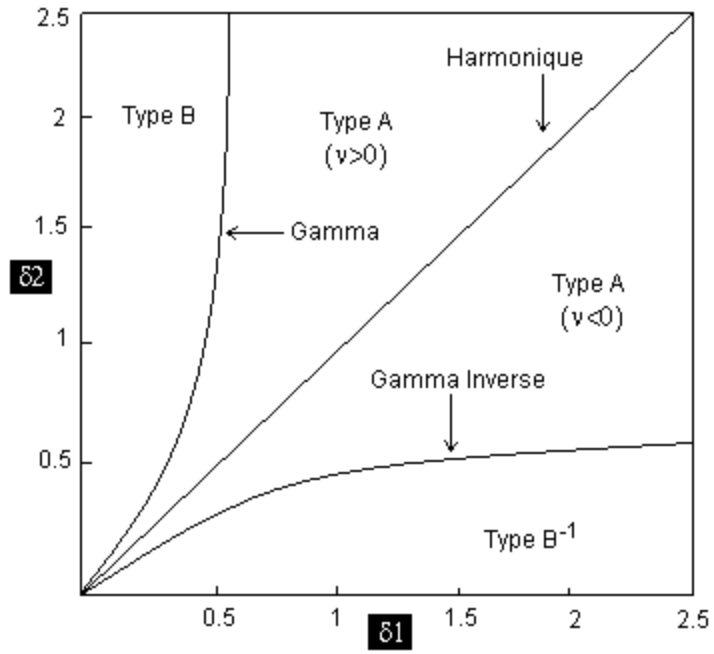
	$n = 50$								$n = 100$							
	RB				RRMSE				RB				RRMSE			
	MM	ML	MMD	MMI	MM	ML	MMD	MMI	MM	ML	MMD	MMI	MM	ML	MMD	MMI
<b>case 1</b>	$C_s = 0.10; C_v = 0.28$															
$T =$	$m = 100; \alpha = 4.00; \nu = 1.20$															
10	<b>0.57</b>	0.25	0.09	0.22	4.19	4.14	4.12	4.16	0.20	0.07	-0.03	0.05	2.96	2.94	<b><u>2.93</u></b>	2.96
100	1.02	1.09	0.48	<b>1.12</b>	5.07	5.07	4.95	5.13	0.46	0.57	0.20	0.56	3.67	3.66	<b><u>3.62</u></b>	3.71
200	1.14	1.29	0.58	<b>1.33</b>	5.31	5.32	5.16	5.43	0.55	0.69	0.26	0.70	3.90	3.85	<b><u>3.80</u></b>	3.96
<b>case 2</b>	$C_s = 0.32; C_v = 0.43$															
$T =$	$m = 100; \alpha = 2.00; \nu = 0.90$															
10	-0.56	-0.71	-0.77	-0.70	5.97	5.91	5.94	5.92	-0.52	-0.60	-0.60	-0.61	4.25	4.21	4.23	4.21
100	-1.80	-1.47	-1.64	-1.43	7.20	6.89	7.10	6.93	-1.30	-1.18	-1.13	-1.16	5.31	5.07	5.18	5.09
200	-2.00	-1.62	-1.79	-1.57	7.50	7.14	7.38	7.18	-1.44	-1.29	-1.22	-1.21	5.57	5.29	5.41	6.32
<b>case 3</b>	$C_s = 0.58; C_v = 0.56$															
$T =$	$m = 100; \alpha = 1.00; \nu = 0.70$															
10	-1.72	-2.01	-1.78	-2.02	7.67	7.66	7.63	7.67	-1.11	-1.29	-1.10	-1.30	5.42	5.43	5.39	5.44
100	-4.90	-5.18	-4.03	-5.17	9.60	9.42	9.08	9.45	-3.13	-3.55	-2.51	-3.50	6.79	6.80	6.42	6.83
200	-5.38	-5.63	-3.80	-5.63	11.65	11.90	19.81	12.43	-3.23	-3.98	-2.59	-3.81	12.91	7.15	11.25	10.23
<b>case 4</b>	$C_s = 0.76; C_v = 0.65$															
$T =$	$m = 100; \alpha = 0.50; \nu = 0.60$															
10	-2.68	-3.08	-2.72	<b><u>-3.09</u></b>	8.91	8.96	8.88	<b>8.97</b>	-1.73	-1.96	-1.71	-2.01	6.63	6.65	6.59	6.70
100	-8.00	-8.84	-6.85	<b><u>-8.85</u></b>	11.94	12.22	11.11	<b>12.25</b>	-5.60	-6.52	-4.84	-6.45	8.85	9.26	8.23	9.32
200	-6.86	-9.86	-7.11	<b><u>-9.87</u></b>	<b>29.45</b>	12.99	18.05	13.03	-5.92	-6.63	-3.62	-6.69	14.84	18.75	23.97	15.71

Numbers underlined and in bold characters represent the minimum value of the criterion (RB or RRMSE) for given  $T$ . Numbers in bold and italic characters represent the maximum value of the criterion (RB or RRMSE) for given  $T$ .

**Tab. 3: RB and RRMSE of quantile estimation of HIB distribution using the four methods (%)**

	$n = 50$								$n = 100$							
	RB				RRMSE				RB				RRMSE			
	MM	ML	MMD	MMI	MM	ML	MMD	MMI	MM	ML	MMD	MMI	MM	ML	MMD	MMI
<b>case 1</b>	$C_s = 1.09; C_v = 0.17$															
$T =$	$m = 100; \alpha = 6.00; \nu = 4.00$															
10	0.47	0.76	0.86	1.05	4.08	4.04	4.07	7.30	0.20	0.31	0.38	0.66	2.66	<u><b>2.64</b></u>	2.66	5.39
100	1.42	0.64	1.68	1.32	7.60	7.02	7.46	<b>17.60</b>	0.61	<u><b>-0.01</b></u>	0.73	0.78	4.86	<u><b>4.58</b></u>	4.80	12.51
200	1.75	0.52	1.97	1.33	8.83	7.98	8.64	21.01	0.76	<u><b>-0.17</b></u>	0.85	0.83	5.64	<u><b>5.23</b></u>	5.55	16.14
<b>case 2</b>	$C_s = 1.42; C_v = 0.22$															
$T =$	$m = 100; \alpha = 4.0; \nu = 3.60$															
10	1.12	1.27	<b>1.42</b>	1.20	5.00	4.96	5.04	5.00	0.14	0.19	0.28	0.18	3.32	3.30	3.32	3.31
100	3.75	2.04	3.83	1.84	10.49	8.91	10.29	9.01	1.40	0.36	1.44	0.36	6.75	6.11	6.64	6.21
200	4.90	2.27	4.88	2.05	12.84	10.29	12.49	10.44	2.03	0.45	2.03	0.48	8.23	7.13	8.05	7.31
<b>case 3</b>	$C_s = 1.80; C_v = 0.26$															
$T =$	$m = 100; \alpha = 3.20; \nu = 3.00$															
10	0.87	0.96	1.12	0.84	5.98	5.91	5.96	5.95	<u><b>0.08</b></u>	0.12	0.22	0.10	4.07	4.06	4.08	4.07
100	4.76	2.00	4.62	1.68	14.47	11.65	13.71	12.01	2.21	0.70	2.22	0.71	8.93	7.97	8.79	8.15
200	7.15	2.46	6.60	2.14	21.27	13.85	18.20	14.47	3.40	1.03	3.35	1.10	11.33	9.51	11.08	9.85
<b>case 4</b>	$C_s = 2.38; C_v = 0.31$															
$T =$	$m = 100; \alpha = 3.00; \nu = 2.40$															
10	0.92	0.98	1.18	0.84	6.73	6.72	<b>6.79</b>	6.73	0.43	0.46	0.56	0.41	4.56	4.54	4.58	4.53
100	<b>6.00</b>	2.08	5.75	1.83	17.11	13.16	16.31	13.52	3.76	1.42	3.69	1.48	11.61	9.31	11.23	9.65
200	<b>9.57</b>	2.57	8.69	2.40	<b>25.84</b>	15.67	22.54	16.40	6.01	1.97	5.76	2.19	16.36	11.33	15.41	12.01

Numbers underlined and in bold characters represent the minimum value of the criterion (RB or RRMSE) for given  $T$ . Numbers in bold and italic characters represent the maximum value of the criterion (RB or RRMSE) for given  $T$ .



**Fig. 1: Halphen distributions in the  $(\delta_1, \delta_2)$  diagram with  $\delta_1 = \ln\left(\frac{A}{G}\right)$  and  $\delta_2 = \ln\left(\frac{G}{H}\right)$**

**(From Morlat, 1956)**

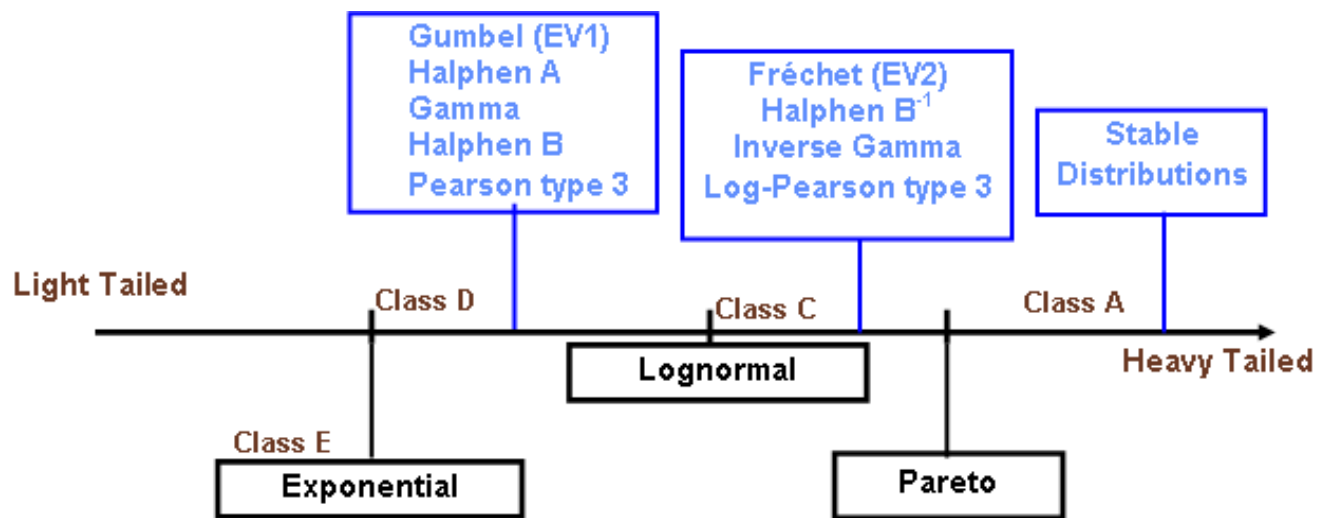
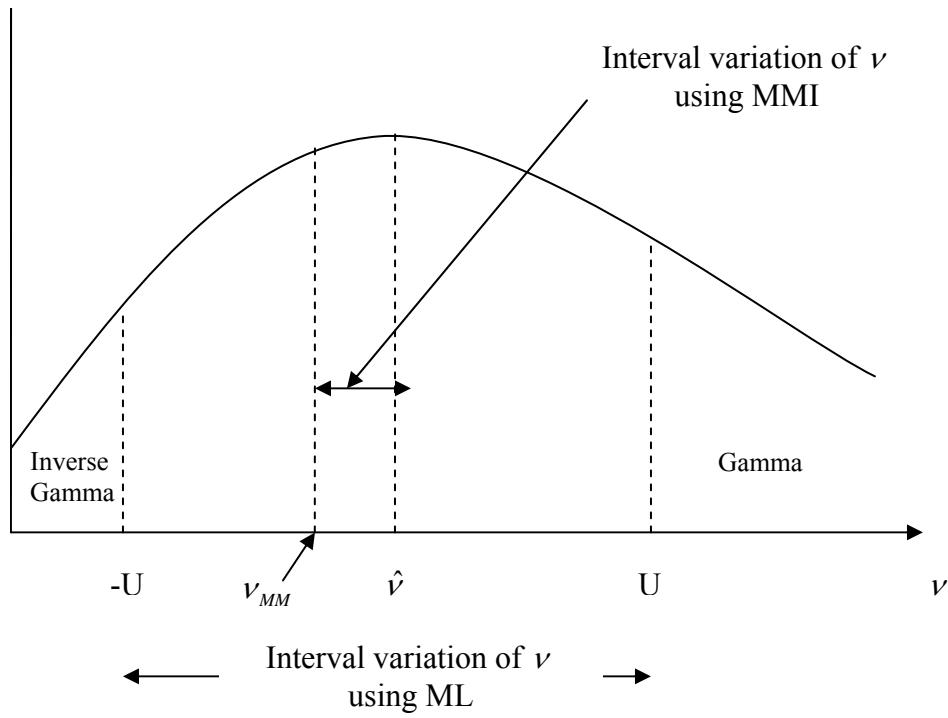
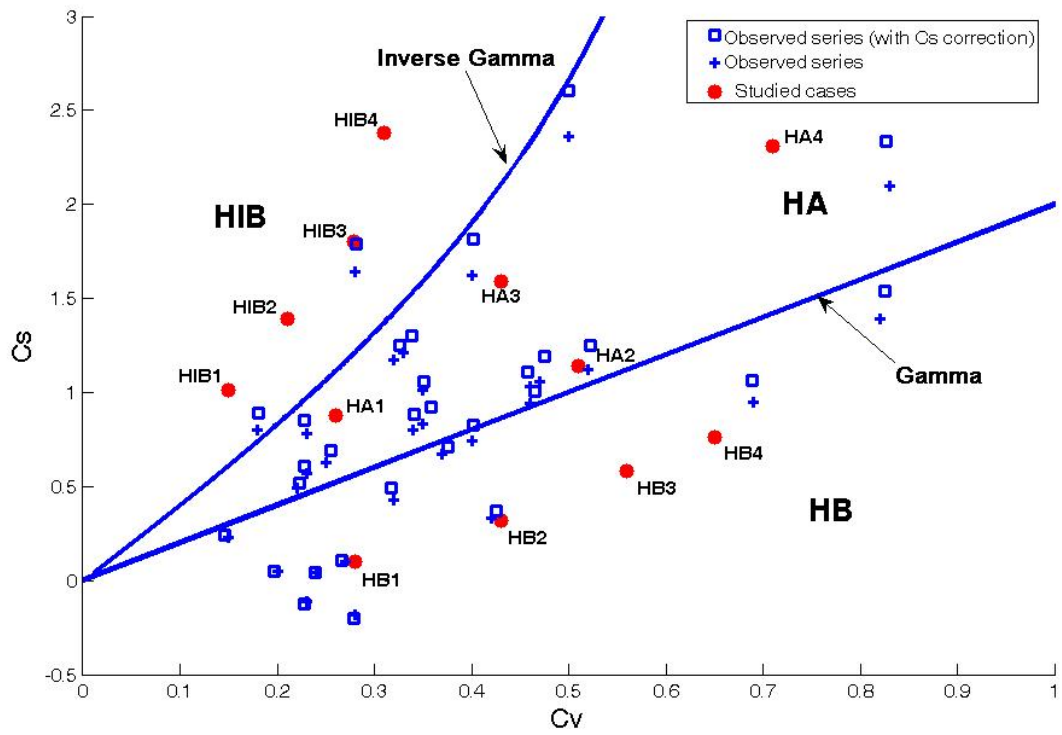


Fig. 2: Tail behaviour classification of distributions commonly used in hydrology (From El Adlouni et al., 2008).

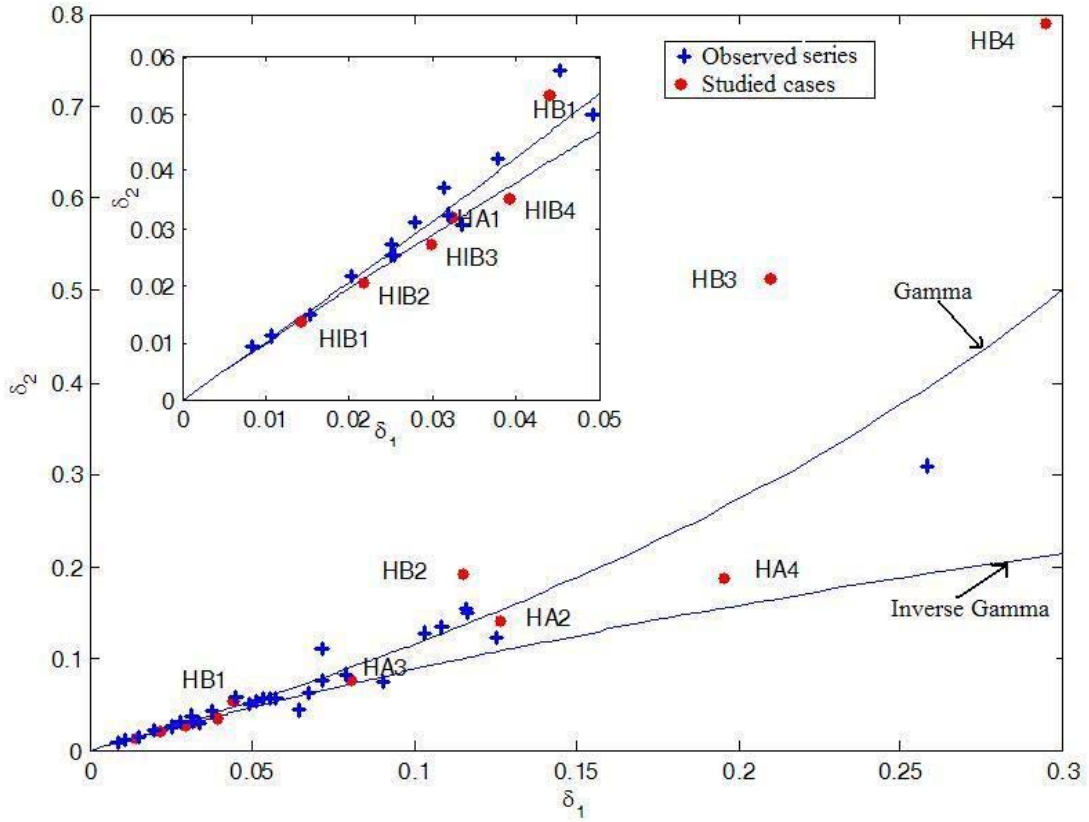
$$\ln L_{HA}(\nu | \alpha, m)$$



**Fig. 3: Illustration of interval variations of  $\nu$  using ML and MMI for HA distribution**



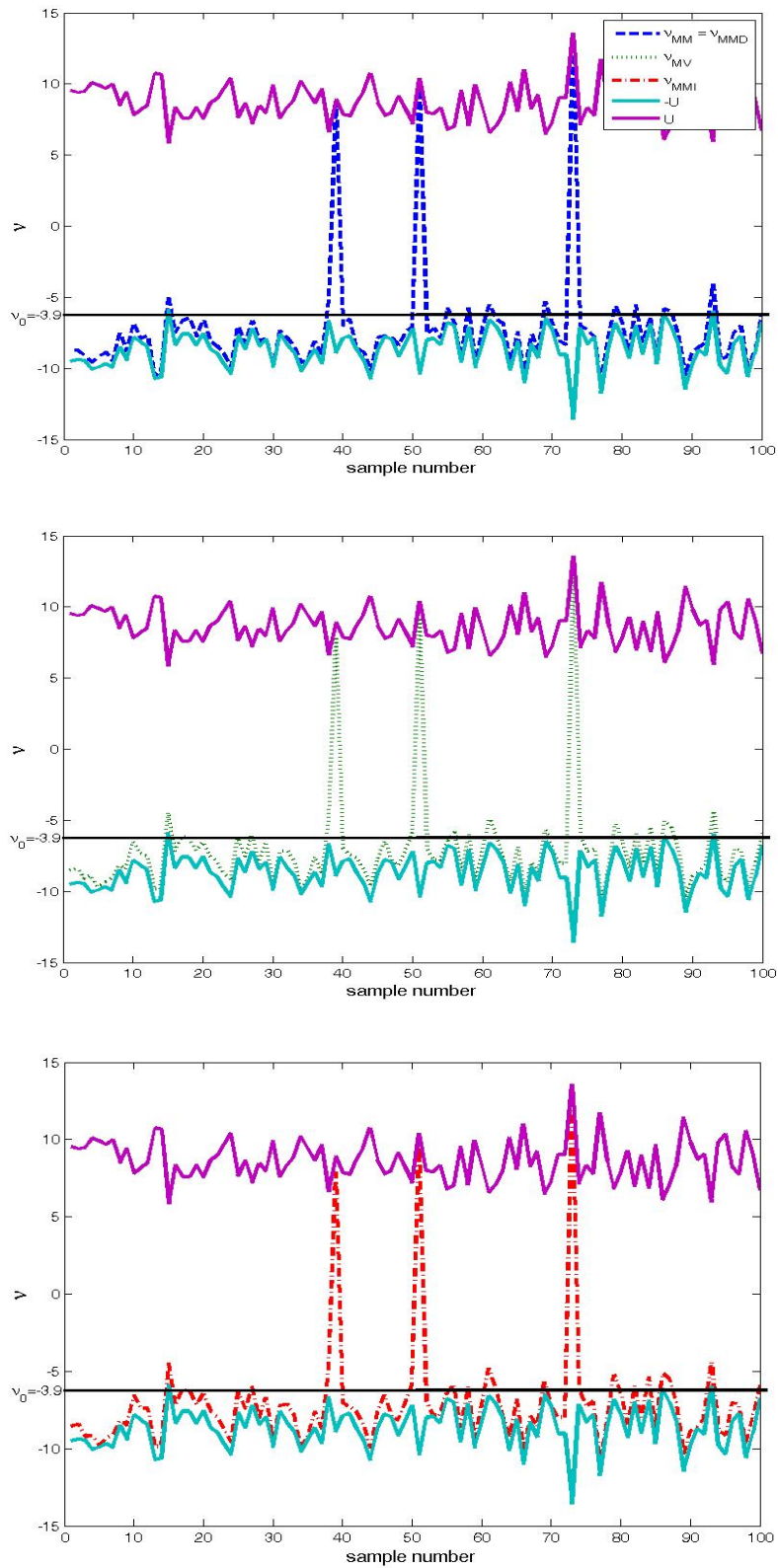
**Fig. 4:** Halphen distributions in the  $C_v$ - $C_s$  diagram with the studied cases and the UNESCO series



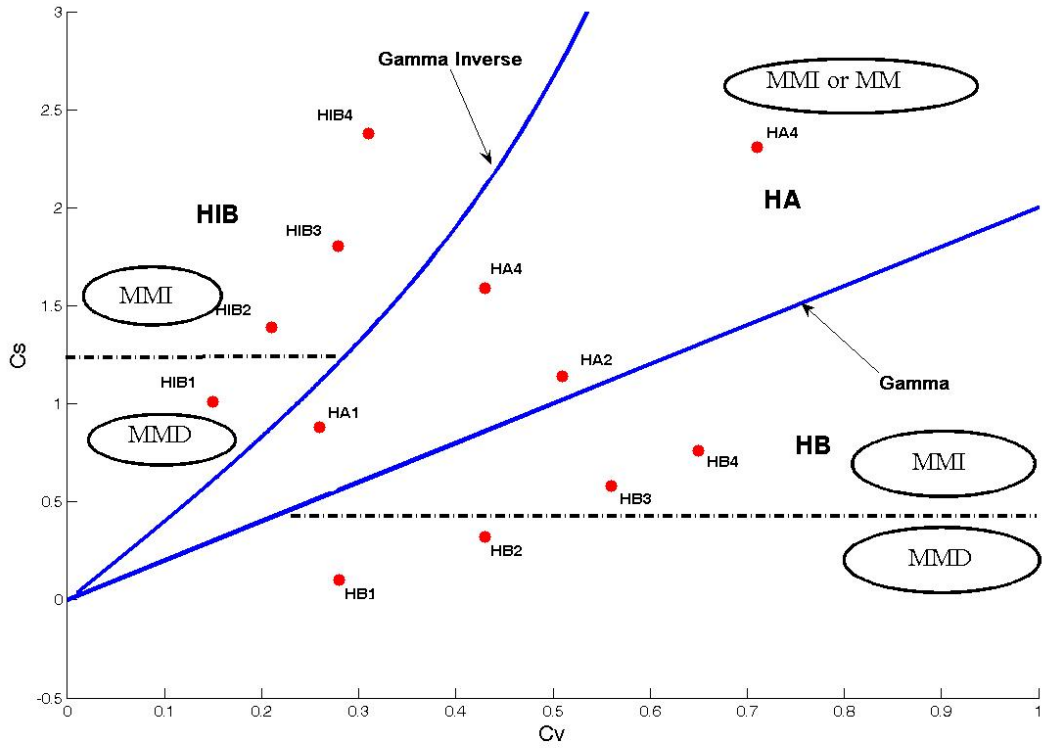
**Fig.5:** The studied cases and the UNESCO series in the  $(\delta_1, \delta_2)$  diagram with  $\delta_1 = \ln\left(\frac{A}{G}\right)$  and

$$\delta_2 = \ln\left(\frac{G}{H}\right)$$

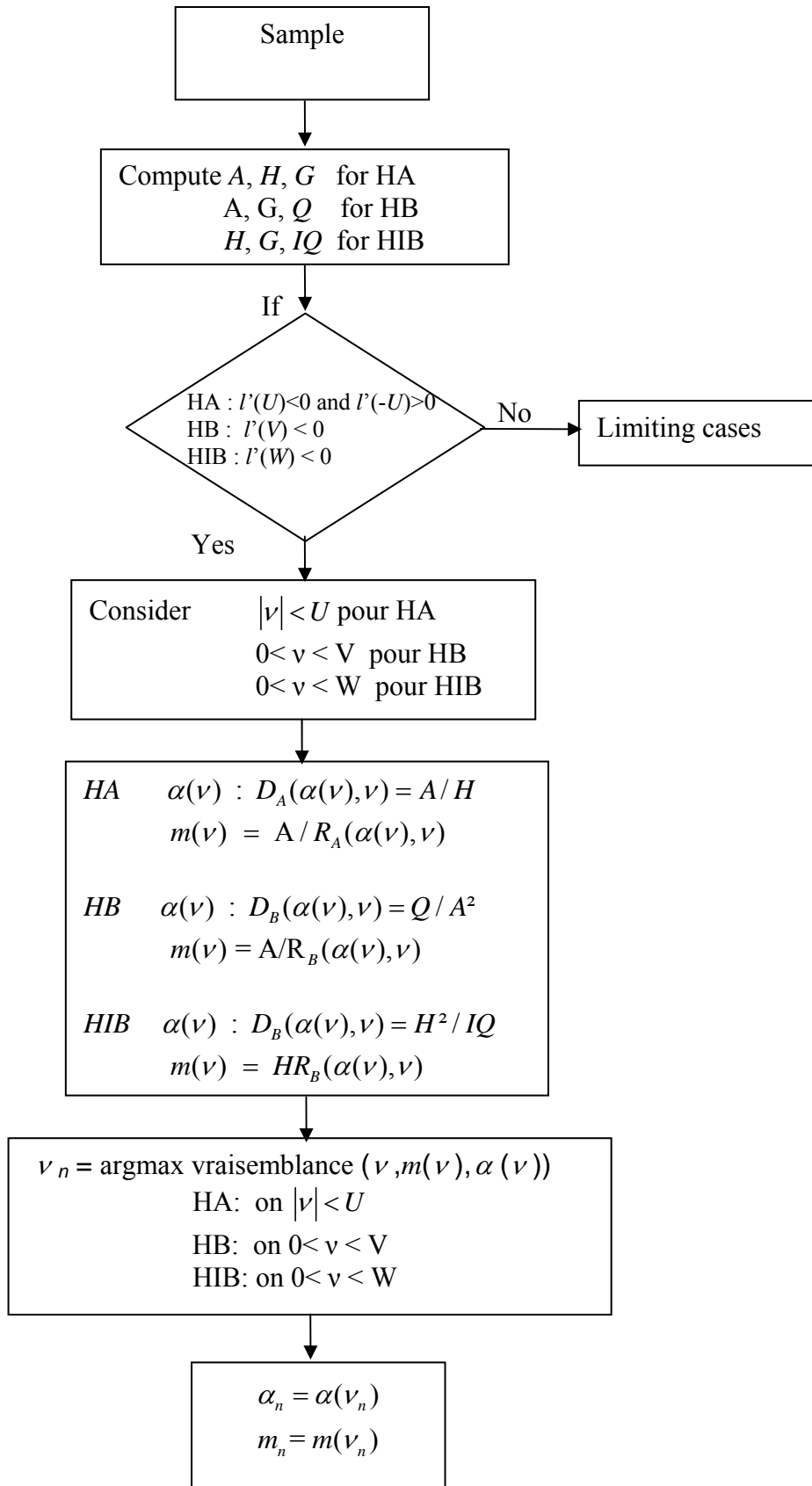




**Fig. 6: Estimation of  $\nu$  and the corresponding empirical variation interval for HA (case 3 and  $n=100$ )**



**Fig. 7: Summary of conclusions related to quantile estimation results**



**Fig. A.1: ML estimation algorithm for parameters of Halphen distributions**