REVSTAT – Statistical Journal
Supplementary file

# Supplementary — Inferences on the Association Parameter in FGM Copula Based Bivariate Distribution: An Application to Water Quality Data

Authors: Rahul Chatterjee
– Department of Mathematics,
University of Louisiana at Lafayette, USA
chatterjee.rahul@epa.gov; rxc9305@gmail.com

Nabendu Pal ✉
– Faculty of Mathematics and Statistics,
Ton Duc Thang University, Vietnam
nabendu.pal@tdtu.edu.vn
and
Department of Mathematics,
University of Louisiana at Lafayette, USA
nabendu.pal@louisiana.edu

## 1. Proof of Lemma 1.1:

Fisher information is this context is the amount of information a data

$$i(\lambda) = \mathrm{E}[\frac{\partial}{\partial\lambda}(lnf(x_1,x_2))]^2,$$

where $f(x_1,x_2)$ is as given in (1.5).

$$\begin{aligned}\frac{\partial}{\partial\lambda}lnf &= (1/f)\frac{\partial}{\partial\lambda}f\\ &= (2F_1(x_1)-1)(2F_2(x_2)-1)(1+\lambda(2F_1(x_1)-1)(2F_2(x_2)-1))^{-1}\end{aligned}$$

Hence the $FIPO$ is

$$\begin{aligned}i(\lambda) &= E[((2F_1(x_1)-1)(2F_2(x_2)-1)/\{1+\lambda(2F_1(x_1)-1)(2F_2(x_2)-1)\})^2]\\ &= \int_{\mathbb{R}_1}\int_{\mathbb{R}_2}\frac{(2F_1(x_1)-1)^2(2F_2(x_2)-1)^2 f_1(x_1)f_2(x_2)}{\{1+\lambda(2F_1(x_1)-1)(2F_2(x_2)-1)\}}\,dx_1\,dx_2\end{aligned}$$

Let $(2(F_i(x_i)-1) = u_i$ then $\frac{du_i}{dx_i} = 2f_i(x_i) \iff f(x_i)dx_i = du_i/2$.

$$i(\lambda) = (1/4)\int_{-1}^{+1}\int_{-1}^{+1} u_1^2u_2^2/(1+\lambda u_1u_2)du_1du_2.$$

✉ Corresponding author

The above double integration can be done symbolically. Another option is to see that both $u_1$ and $u_2$ are between $-1$ and $+1$, i.e., $|u_i| < 1$. Since $|\lambda| < 1$, $|\lambda u_1 u_2| < 1$. So, we can use the series expression:

$$(1 + \lambda u_1 u_2)^{-1} = \sum_{j=0}^{\infty} (-1)^j (\lambda u_1 u_2)^j.$$

Hence,

$$\begin{aligned} i(\lambda) &= (1/4) \int_{-1}^{+1} \int_{-1}^{+1} u_1^2 u_2^2 / (1 + \lambda u_1 u_2) du_1 du_2 \\ &= (1/4) \sum_{j=0}^{\infty} (-1)^j (\lambda)^j \Big( \int_{-1}^{+1} u_1^{j+2} du_1 \Big) \Big( \int_{-1}^{+1} u_2^{j+2} du_2 \Big) \\ &= (1/4) \sum_{j=0}^{\infty} (-1)^j (\lambda)^j \{ u^{j+3}/(j+3) |_{-1}^{+1} \}^2 \\ &= (1/4) \sum_{j=0}^{\infty} (-1)^j (\lambda)^j / (j+3) \{ (1)^{j+3} - (-1)^{j+3} \}^2. \end{aligned}$$

In the above expression all the odd terms would vanish, and the even terms would remain. Let $j = 2m$, $m = 0, 1, 2, ....$ So,

$$i(\lambda) = \lambda^{2m}/(2m+3).$$

It is worthy to note here that the infinite sum in the above expression is convergent and has a finite expression. The final form of the $FIPO$ is given by

$$i(\lambda) = \{-\lambda + tanh^{-1}(\lambda)\}/\lambda^3, \tag{1.1}$$

where $tanh^{-1}(\lambda) = \frac{1}{2} log((1+\lambda)/(1-\lambda))$.

## 2. Proof of Theorem 2.1

A direct differentiation of (2.3) and then maximizing it by equating the derivative to zero is not the right approach to find $\hat{\lambda}_{ML}$ numerically.

First define $a_i$, $1 \leq i \leq n$ as $a_i = G(x_{i1})G(x_{i2})$. It is easy to see that each $a_i \in [-1, +1]$, since each $G(x_{ik})$, $k = 1, 2$, is so. Further note that each $a_i > 0$ if and only if $G(x_{ik}) \gtrless 0$ for both $k = 1, 2$. Similarly, each $a_i < 0$ if $G(x_{i1}) > 0$ and $G(x_{i2}) < 0$, or $G(x_{i1}) < 0$ and $G(x_{i2}) > 0$. let $M_k$ be the median of the marginal $f_k$, $k = 1, 2$. Hence, $a_i > 0$ provided either both $x_{ik} > M_k$ $(k = 1, 2)$ or both $x_{ik} > M_k$ $(k = 1, 2)$. Thus, each $a_i > 0$ with probability 0.5, and $< 0$ with probability 0.5.

The derivative of $l(\lambda)$ w.r.t $\lambda$ is $h(\lambda)$, where $h(\lambda) = \sum_{i=1}^{n} a_i/(1+\lambda a_i)$. From a numerical point of view it is not advisable to find $\hat{\lambda}_{ML}$ by setting $h(\lambda) = 0$, since there may not be such a solution. Hence, one needs to proceeds carefully.

Note that $h'(\lambda) = -\sum_{i=1}^{n}\{a_i/(1+\lambda a_i)\}^2 < 0 \,\forall\, \lambda$, i.e., $h(\lambda)$ is monotonically decreasing. Two expressions are very important here: $h(-1) = \sum_{i=1}^{n} a_i/(1-a_i)$, and $h(+1) = \sum_{i=1}^{n} a_i/(1+a_i)$. Now we are going to consider three cases as given below.

(a) $h(-1) > 0$ and $h(+1) < 0$.

In this case, $h(\lambda)$ is decreasing monotonically, starting with some positive value at $\lambda = -1$ and then ending with some negative value at $\lambda = +1$. So there exists a unique $\lambda_0 \in (-1,+1) \ni h(\lambda_0) = 0$. this $\lambda_0$ is the value of $\hat{\lambda}_{ML}$, and it is unique by the monotonicity of $h(\lambda)$.

(b) $h(-1) < 0$ ( which also implies that $h(+1) < 0$).

Since $h(\lambda)$ is monotonically decreasing, $h(\lambda)$ is negative over the entire space of $\lambda \in [-1,+1]$. Since $h(\lambda) = l'(\lambda)$, it implies that when our observations are such that $h(-1) = \sum_{i=1}^{n} a_i/(1-a_i) < 0$, the log-likelihood function is monotonically decreasing for $\lambda \in [-1,+1]$, i.e., $l(\lambda)$ is attaining its supremum at $\hat{\lambda}_{ML} = -1$, and this is unique.

(c) $h(+1) > 0$ ( which also implies that $h(-1) > 0$).

Since $h(\lambda)$ is monotonically decreasing, $h(\lambda)$ is positive over the entire space of $\lambda \in [-1,+1]$. Since $h(\lambda) = l'(\lambda)$, this implies that when our observations are such that $h(+1) = \sum_{i=1}^{n} a_i/(1+a_i) > 0$, the log-likelihood function is monotonically decreasing for $\lambda \in [-1,+1]$, i.e., $l(\lambda)$ is attaining its supremum at $\hat{\lambda}_{ML} = +1$, and this is unique.

## 3. Bayes' estimator under the approximate Jeffrey's prior ($\hat{\lambda}_{BAJP}$)

Equation (2.14) gives the general form of a Bayes' estimate of $\lambda$ with a prior distribution $\pi(\lambda)$. The algebraic restructuring in (2.15) gives us a more tractable form the general Bayes' estimator which is given by (2.16). In (2.16) call the numerator as $A$ and the denominator as $B$. We are going to use $w(\lambda) = 1$ and $\pi(\lambda) = \pi_{AJP}(\lambda)$, where $\pi_{AJP}(\lambda)$ is given by (2.21). Below we derive both $A$ and $B$ separately to arrive at the final form of $\hat{\lambda}_{BAJP}$ given in (2.22).

$$
\begin{aligned}
A &= \int_{-1}^{1} \lambda w(\lambda) \sum_{k=0}^{n} D_k \lambda^k \sum_{m=0}^{\infty} |\lambda|^m (2m+3)^{-1/2} d\lambda \\
&= \sum_{k=0}^{n}\sum_{m=0}^{\infty} D_k (2m+3)^{-1/2} \int_{-1}^{1} \lambda^{k+1} w(\lambda) |\lambda|^m d\lambda \\
&= \sum_{k=0}^{n}\sum_{m=0}^{\infty} D_k (2m+3)^{-1/2} [\int_{0}^{1} \lambda^{k+1} (\lambda)^m d\lambda + \int_{-1}^{0} \lambda^{k+1} (-\lambda)^m d\lambda].
\end{aligned}
\tag{3.1}
$$

Let us focus on the integration term within the square bracket in (D.1). Call $\int_0^1 \lambda^{k+1}(\lambda)^m d\lambda + \int_{-1}^{0} \lambda^{k+1}(-\lambda)^m d\lambda = c_1$. In $c_1$, consider the second part of the integration i.e., $\int_{-1}^{0} \lambda^{k+1}(-\lambda)^m d\lambda$. In this integration let $-\lambda = u \implies -d\lambda = du$ and with this substitution the integration becomes $\int_{+1}^{0}(-u)^{k+1}u^m(-du) = (-1)^{k+1}\int_0^1 u^{m+k+1}du$. So,

$$\begin{aligned} c_1 = &\int_0^1 \lambda^{m+k+1}d\lambda + (-1)^{k+1}\int_0^1 \lambda^{m+k+1}d\lambda \\ =&\{1+(-1)^{k+1}\}\int_0^1 \lambda^{m+k+1}d\lambda \\ =&\{1+(-1)^{k+1}\}(m+k+2)^{-1}. \end{aligned}$$

So, $A = \sum_{k=0}^{n}\sum_{m=0}^{\infty} D_k(2m+3)^{(-1/2)}(1+(-1)^{k+1})(m+k+2)^{-1}$.
Similarly, $B = \sum_{k=0}^{n}\sum_{m=0}^{\infty} D_k(2m+3)^{-1/2}[\int_0^1 \lambda^k(\lambda)^m d\lambda + \int_{-1}^{0}\lambda^k(-\lambda)^m d\lambda]$ i.e., $B = \sum_{k=0}^{n}\sum_{m=0}^{\infty} D_k(2m+3)^{(-1/2)}(1+(-1)^k)(m+k+1)^{-1}$. So, $\hat{\lambda}_{BAJP} = A/B$, which is given in (2.22).

## 4. Application Data

**Table 1**: MDR groundwater Data

| Well ID | As (ppb) | Cl (ppm) | Eh (mv) | pH | Well ID | As (ppb) | Cl (ppm) | Eh (mv) | pH |
|---|---|---|---|---|---|---|---|---|---|
| DT7 | 563.9 | 107 | -126 | 6.78 | TB19 | 300.3 | 160.3 | -120 | 6.68 |
| DT6 | 0.5 | 56.1 | 142 | 6.71 | TBE10 | 700.4 | 81.4 | -108 | 7.07 |
| DT5 | 0.7 | 46.8 | 199 | 7.04 | TBE9 | 196.2 | 986.6 | -110 | 6.72 |
| DT3 | 0.4 | 345.2 | 169 | 6.44 | TBE7 | 166.3 | 20 | -84 | 7.16 |
| DT4 | 0.1 | 500.1 | 165 | 6.51 | TBE4 | 4.4 | 1499.6 | 82 | 6.09 |
| DT2 | 1.8 | 632.8 | 101 | 6.66 | TBE5 | 981.4 | 60.4 | -110 | 6.87 |
| DT1 | 13.1 | 19.7 | 97 | 7.75 | TBE3 | 6.8 | 2.7 | 158 | 7.17 |
| TB11 | 462.3 | 9.2 | -114 | 6.92 | TBE1 | 6.6 | 61.7 | 126 | 7.1 |
| TB18 | 155.7 | 25.9 | -72 | 6.52 | TBE11 | 5.3 | 12.2 | 60 | 7.16 |
| TB9 | 187.6 | 12.8 | -128 | 6.94 | TBE6 | 3.2 | 1527 | 149 | 6.73 |
| TB2 | 850.4 | 10.5 | -133 | 7.14 | TH16 | 0.4 | 173.6 | 157 | 6.14 |
| TB24 | 370.4 | 13.9 | -90 | 7.15 | TH9 | 0.2 | 275.3 | 253 | 5.84 |
| TB26 | 139.9 | 13.8 | -83 | 7.43 | TH13 | 0 | 22.7 | 194 | 6.19 |
| TB27 | 77.7 | 5.4 | -33 | 7.24 | TH14 | 0.3 | 113.6 | 184 | 6.02 |
| TB21 | 842.1 | 21.1 | -105 | 6.88 | TH22 | 0.1 | 228.1 | 226 | 6.5 |
| TB1 | 276.8 | 19.6 | -92 | 6.85 | TH21 | 0.3 | 89.9 | 169 | 6.1 |
| TB10 | 377.3 | 8.2 | -129 | 6.79 | TH5 | 0.8 | 742.1 | 251 | 5.83 |
| TB25 | 272.9 | 11.9 | -104 | 7.2 | TH12 | 2.3 | 182.7 | 127 | 6.31 |
| TB13 | 746 | 72.7 | -125 | 7.16 | TH15 | 8.4 | 27.5 | 60 | 6.18 |
| TB22 | 311 | 13.5 | -130 | 6.63 | TH1 | 6 | 544.4 | 210 | 6.08 |
| TB15 | 937.7 | 19.3 | -110 | 7.04 | TH10 | 3.2 | 277.3 | 231 | 6 |
| TB16 | 314.5 | 25.8 | -115 | 6.74 | TH2 | 2 | 487.6 | 130 | 5.87 |
| TB20 | 746.3 | 6.9 | -139 | 6.61 | TH23 | 0.2 | 158.5 | 175 | 6.04 |
| TB23 | 270 | 12.7 | -110 | 7.01 | TH3 | 1.5 | 560.2 | 261 | 6 |
| TB17 | 224.2 | 21.5 | -126 | 6.46 | TH4 | 2.6 | 21.4 | 80 | 5.99 |
| TB3 | 727 | 10.8 | -136 | 7.14 | TH11 | 8.9 | 479.8 | 158 | 6.56 |
| TB12 | 931.5 | 2.9 | -125 | 7.03 | TH18 | 3.6 | 335.6 | 181 | 6.29 |
| TB14 | 747.7 | 63.4 | -115 | 7.15 | TH8 | 6 | 253.2 | 235 | 5.85 |
| TB5 | 416.3 | 0 | -60 | 7.69 | TH7 | 0.7 | 122.3 | 162 | 6.29 |
| TB4 | 360.3 | 42.9 | -130 | 7.34 | TH6 | 0 | 242.8 | 200 | 6.19 |
| TB6 | 315.5 | 61.2 | -111 | 7.36 | TH17 | 22.2 | 40.5 | -13 | 7.03 |
| TB7 | 101.1 | 42.7 | -28 | 7.3 | TH19 | 17.5 | 57.3 | 145 | 7.39 |
| TB8 | 237.6 | 124.4 | -98 | 7.17 | TH20 | 2.4 | 0 | 24 | 6.51 |

**NOTE**: Under the column "Well ID", the wells located in the southern subregion starts with"DT" and "TB". The wells that are located in the northern subregion starts with "TH".

## 5. Relative Frequency Histogram Plots of the Six Estimators

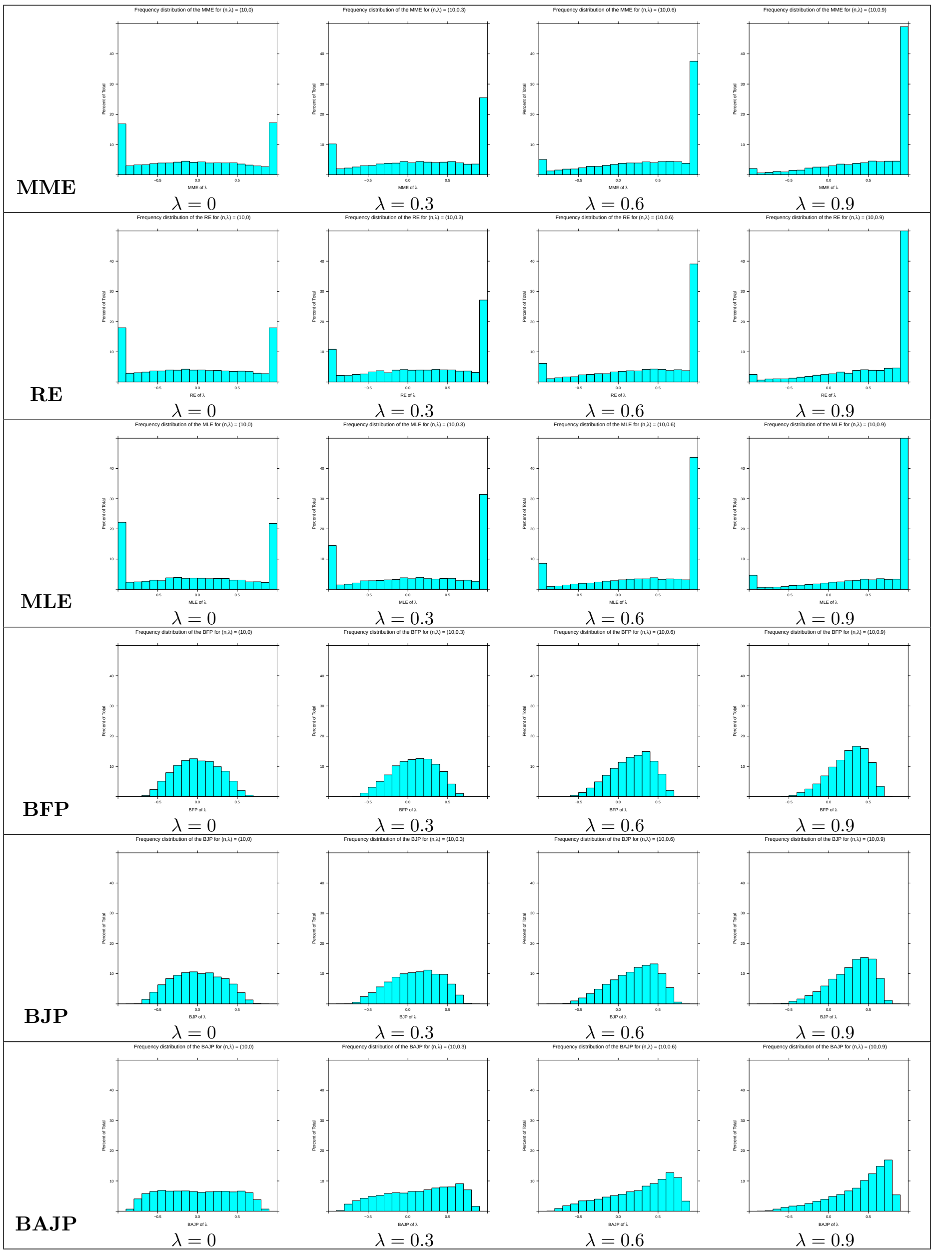



**Figure 1**: Simulated relative frequency histograms of the six estimators of $\lambda$, $n = 10$.

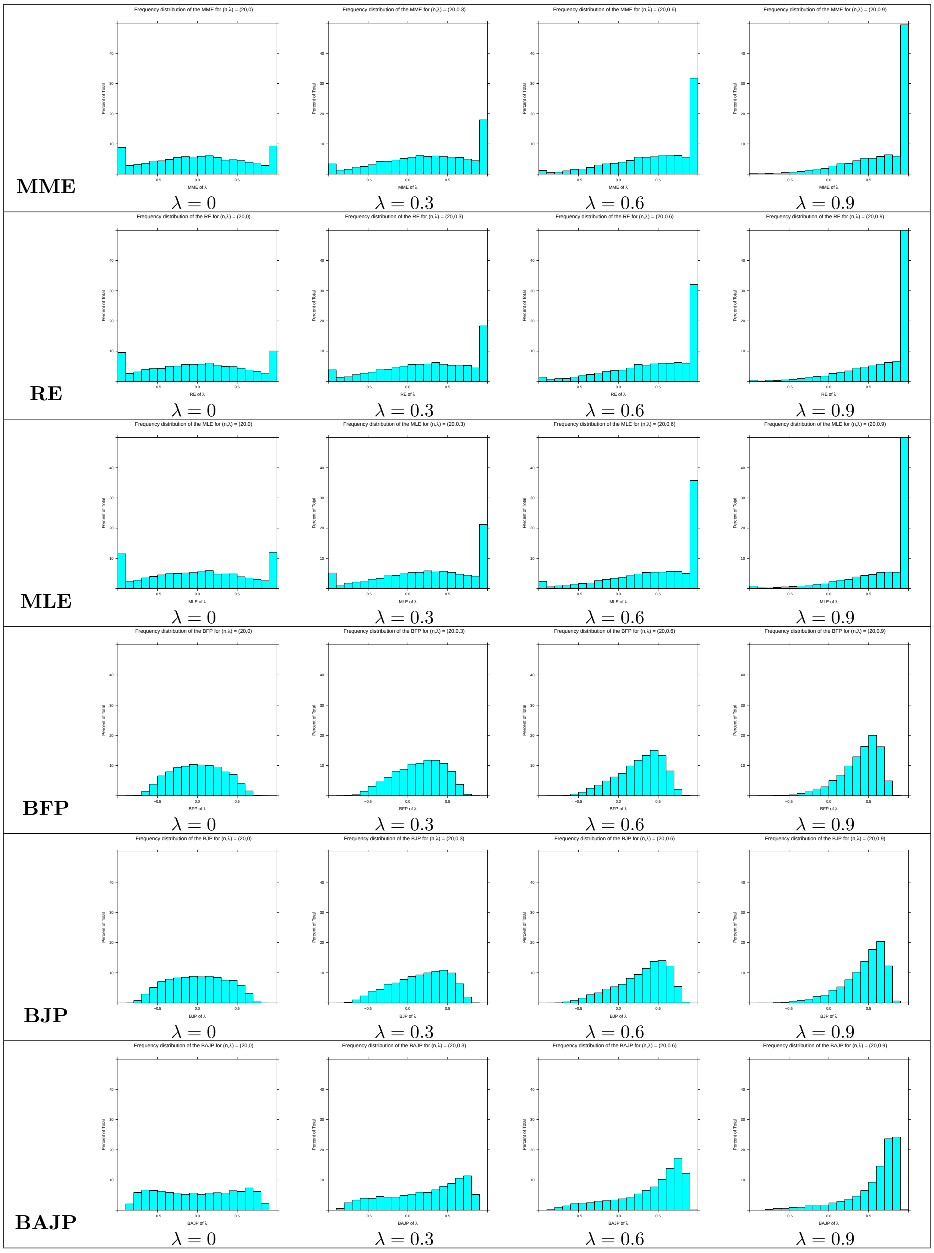


**Figure 2**: Simulated relative frequency histograms of six estimators of $\lambda$, $n = 20$

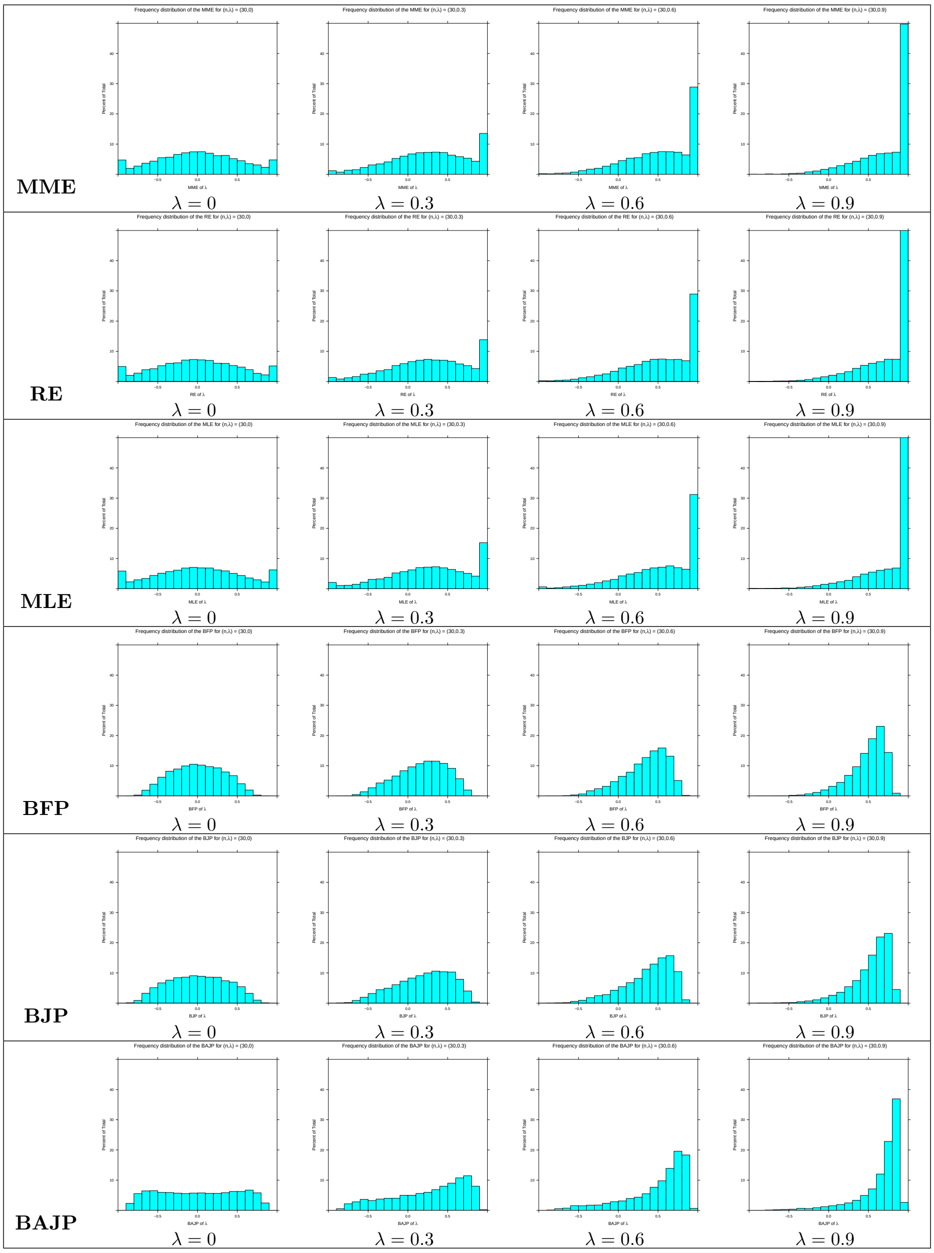



**Figure 3**: Simulated relative frequency histograms of the six estimators of $\lambda$, $n = 30$.

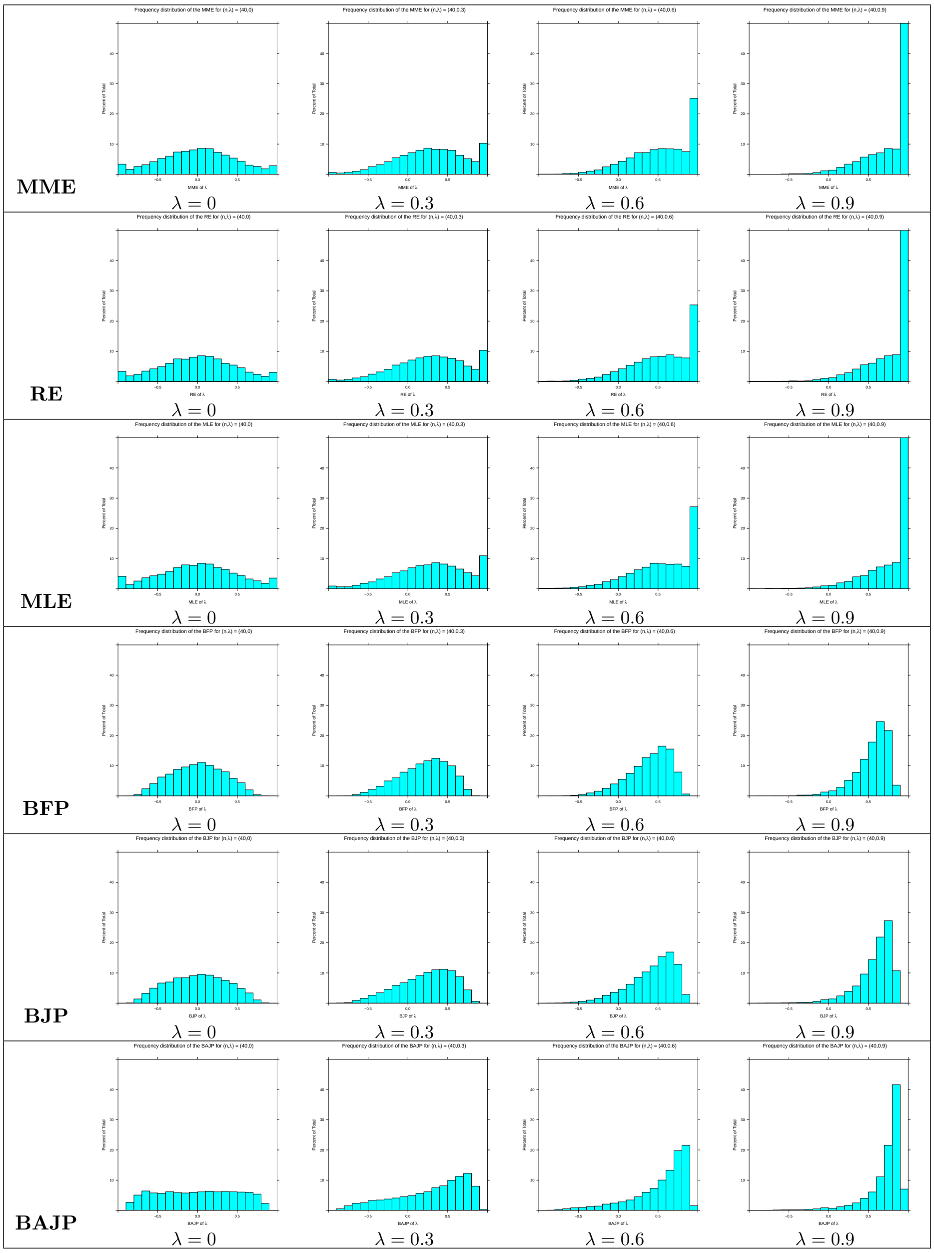



**Figure 4**: Simulated relative frequency histograms of the six estimators of $\lambda$, $n = 40$.

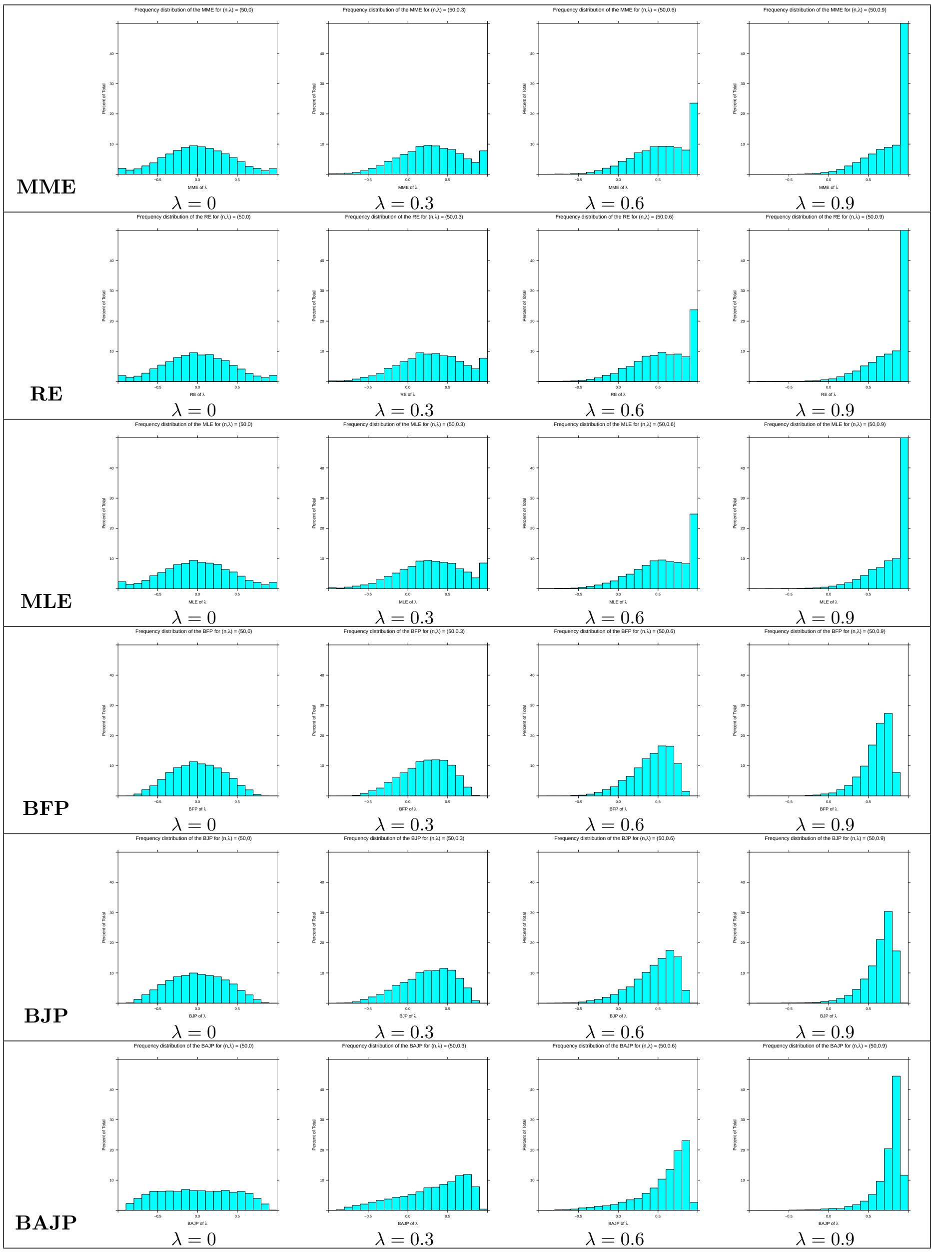


**Figure 5**: Simulated relative frequency histograms of the six estimators of $\lambda$, $n = 50$.

## 6. Bias and MSE Plots of the Six Estimators

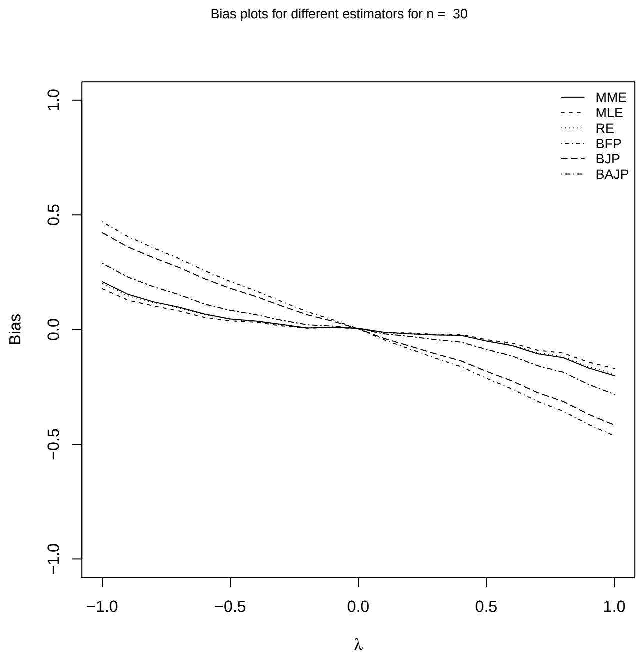



**Figure 6**: Bias plots of the six estimators for $n = 30$

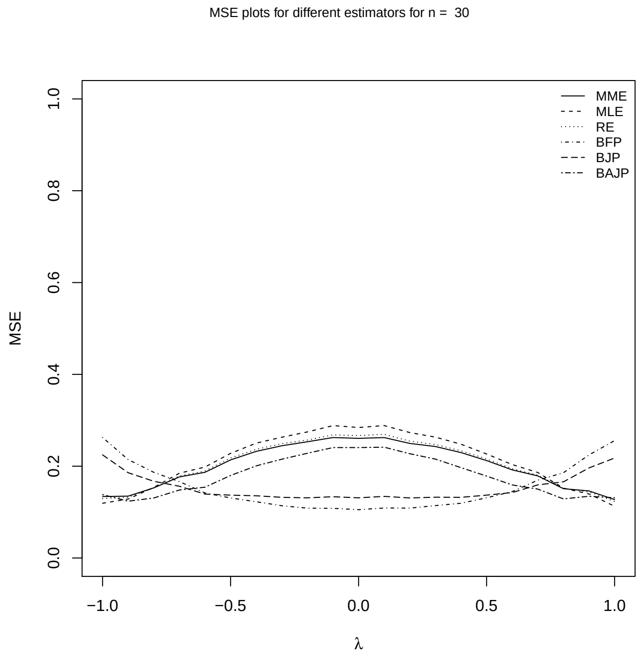
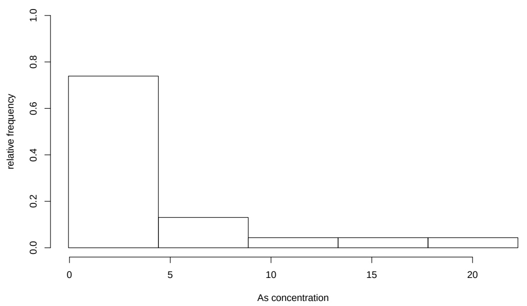
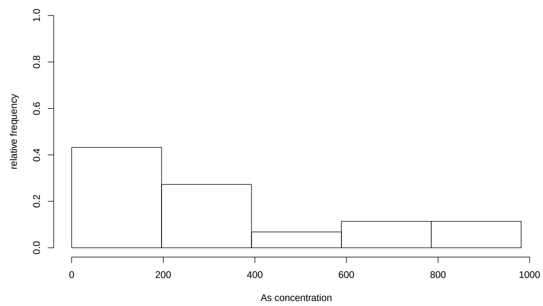



**Figure 7**: MSE plots of the six estimators for $n = 30$

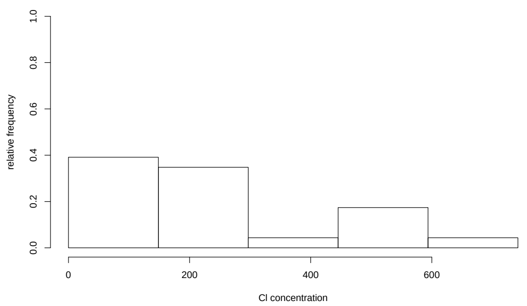
## 7. Relative Frequency Histograms of elements in MDR groundwater



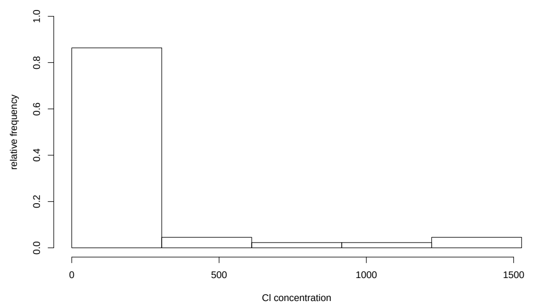

(a) As (North)



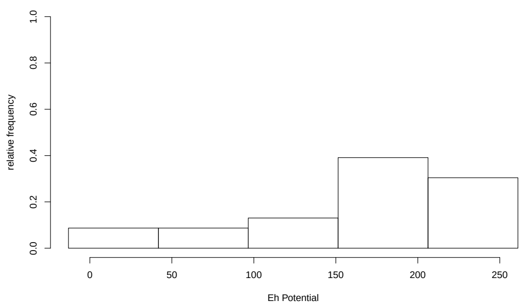

(b) As (South)



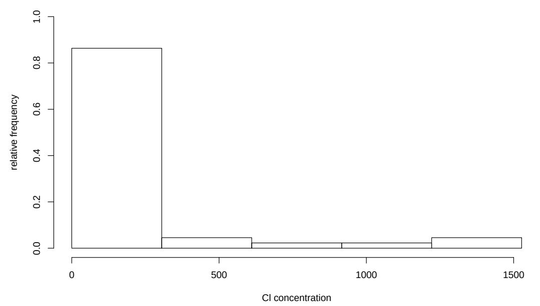

(c) Cl (North)



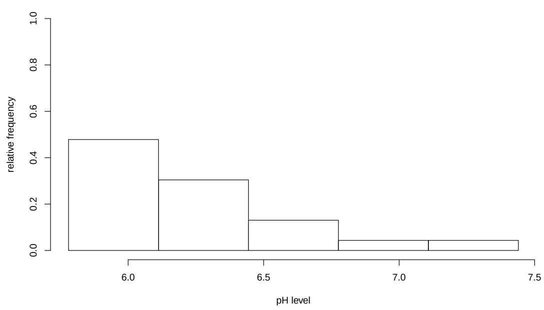

(d) Cl (South)



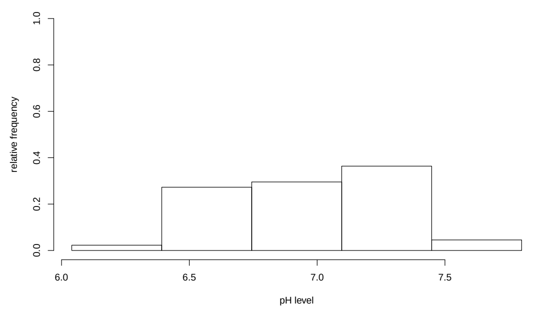

(e) Eh (North)




(f) Eh (South)



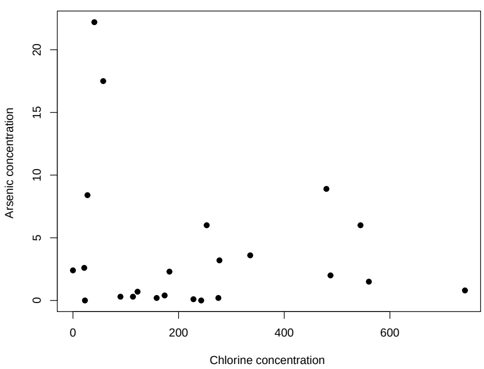

(g) pH (North)



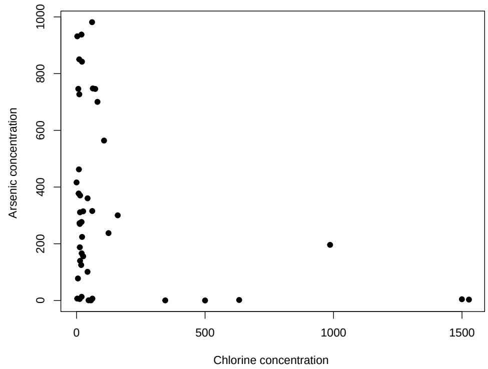

(h) pH (South)

**Figure 8**: Relative frequency histogram of the four elements in two subregions

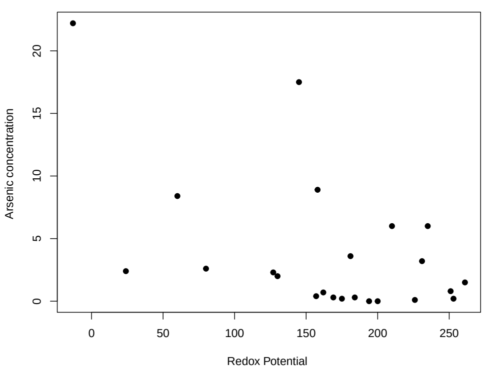
## 8. Scatter plots of Arsenic and the other elements in MDR groundwater



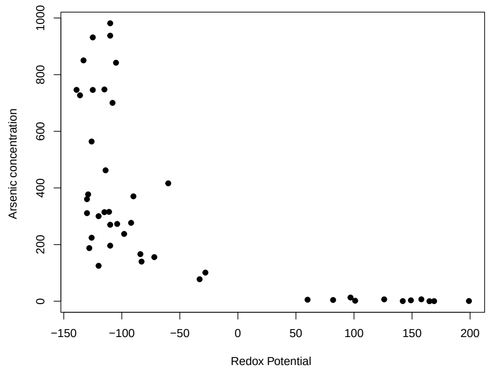

(a) As vs Cl (North)



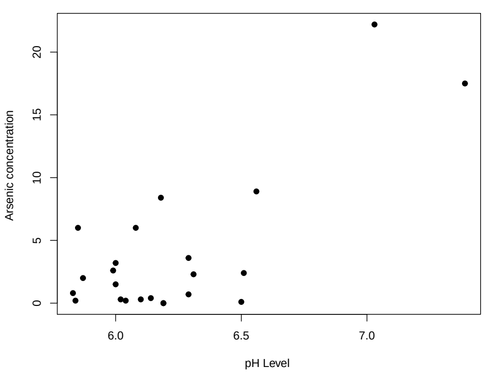

(b) As vs Cl (South)



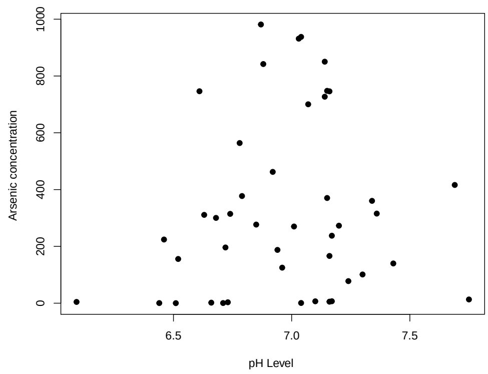

(c) As vs Eh (North)




(d) As vs Eh (South)




(e) As vs pH (North)




(f) As vs pH (South)

**Figure 9**: Scatter plots of As against each of other three elements