

Article

# Bayesian Inference in Extremes Using the Four-Parameter Kappa Distribution

Palakorn Seenoi <sup>1</sup>, Piyapatr Busababodhin <sup>2</sup> and Jeong-Soo Park <sup>3,\*</sup><sup>1</sup> Department of Statistics, Khon Kaen University, Khon Kaen 40002, Thailand; palakorns@kku.ac.th<sup>2</sup> Department of Mathematics, Mahasarakham University, Maha Sarakham 44150, Thailand; piyapatr.b@msu.ac.th<sup>3</sup> Department of Mathematics and Statistics, Chonnam National University, Gwangju 61186, Korea

\* Correspondence: jspark@jnu.ac.kr; Tel.: +82-2-530-3445

Received: 12 November 2020; Accepted: 2 December 2020; Published: 7 December 2020



**Abstract:** Maximum likelihood estimation (MLE) of the four-parameter kappa distribution (K4D) is known to be occasionally unstable for small sample sizes and to be very sensitive to outliers. To overcome this problem, this study proposes Bayesian analysis of the K4D. Bayesian estimators are obtained by virtue of a posterior distribution using the random walk Metropolis–Hastings algorithm. Five different priors are considered. The properties of the Bayesian estimators are verified in a simulation study. The empirical Bayesian method turns out to work well. Our approach is then compared to the MLE and the method of the L-moments estimator by calculating the 20-year return level, the confidence interval, and various goodness-of-fit measures. It is also compared to modeling using the generalized extreme value distribution. We illustrate the usefulness of our approach in an application to the annual maximum wind speeds in Udon Thani, Thailand, and to the annual maximum sea-levels in Fremantle, Australia. In the latter example, non-stationarity is modeled through a trend in time on the location parameter. We conclude that Bayesian inference for K4D may be substantially useful for modeling extreme events.

**Keywords:** deviance information criterion; extreme values; highest posterior density; Markov chain Monte Carlo; profile likelihood; quantile estimation

## 1. Introduction

Introduced by Hosking [1], the four-parameter kappa distribution (K4D) is a generalized form of some frequently used distributions such as the generalized Pareto, logistic, Gumbel, and extreme value (GEV) distributions. The flexibility of the K4D enables its wide application in fitting data when three parameters are not sufficient. The K4D has been used in many fields, including hydrological sciences (e.g., [2–9]). Blum et al. [10] found that the K4D provides a very good representation of daily streamflow across most physiographic regions in the conterminous United States.

The method of L-moments estimation (LME) has been employed for parameter estimation in the K4D as it provides a reliable and robust parameter estimator, particularly for small sample sizes. However, the LME is sometimes non-feasible with the K4D, in that the estimator is not computable within the feasible parameter space [4,11]. Dupuis and Winchester [11] compared the LME and the maximum likelihood estimator (MLE) of the K4D by simulation. The MLE is optimal for large samples and is applicable to various situations, including non-stationary models. Nonetheless, we found in this study that the MLE of the K4D with a small sample produces relatively large variances of the estimated parameters compared to the LME.

The MLE has a larger sampling variance than LME, whereas LME is right biased. The MLE for a generalized extreme value distribution (GEVD) shows a smaller variance, which may indicate

the difficulty of estimating four parameters using the MLE. From these observations, we suggest that a method for a more stable estimation of the parameters may be useful for the K4D. We thus consider a Bayesian estimation because its facility to include other sources of knowledge through a prior distribution has considerable appeal, and it can supply a more comprehensive inference than the corresponding likelihood inference [12], p. 172.

In dealing with extreme events such as annual maximum daily precipitation, seven-thousand three-hundred five observations over 20 years are reduced to only 20 values. This loss of information can often lead to unreliable model estimates, especially for very high quantiles, which may have large variance. Applying Bayesian inference to this problem would permit any additional information about the processes to be used as prior knowledge [13]. Bayesian inference offers an advantage in extreme value analysis with small samples as it allows the incorporation of prior knowledge into the information provided by observed data to improve statistical inference. In addition, Bayesian inference supplies a viable alternative even when the likelihood theory breaks down [12].

Coles and Tawn [14,15] applied Bayesian techniques to model extreme daily rainfall and extreme surges on the east coast of the U.K. Vidal [16] proposed a Bayesian inference method for the Gumbel distribution using three different priors for the scale parameter and a non-informative prior for the location parameter. Lima et al. [17] applied Bayesian analysis to the estimation of intensity–duration–frequency curves for rainfall data. Lima et al. [18] proposed a hierarchical Bayesian GEV model to improve local and regional flood estimations. Fawcett and Walshaw [19] demonstrated predictive inference for return values of sea surge and wind speed extremes, and Fawcett and Green [20] investigated posterior predictive return levels for wind speed extremes in the U.K. However, Bayesian inference for the K4D has not been considered yet, which motivates this study. This is the first time Bayesian inference for the K4D with a non-stationary example has been studied. We employ five priors: one is non-informative, and four are informative priors. An empirical Bayes method is considered by using the data to get information on the prior parameters.

The rest of this paper is organized as follows. The K4D is introduced in Section 2. The MLE and L-moments are presented in Section 3. In Section 4, we propose a Bayesian procedure for parameter estimation of the K4D using a random walk Metropolis–Hastings algorithm. Numerical comparisons among the MLE, LME, and Bayesian estimators (BEs) are investigated via simulation in Section 5. Section 6 presents applications to real-world data. Finally, concluding remarks are presented in Section 7. The technical specifics and some figures are provided in the accompanying Supplementary Material (SM).

## 2. Four-Parameter Kappa Distribution

Let  $X$  be a random variable of the K4D introduced by Hosking [1]. The cumulative distribution function (cdf) of  $X$  for  $k \neq 0$  and  $h \neq 0$  is expressed as:

$$F(x) = \left\{ 1 - h [1 - k(x - \zeta)/\alpha]^{1/k} \right\}^{1/h}, \tag{1}$$

where  $k$  and  $h$  are shape parameters,  $\zeta$  is a location parameter, and  $\alpha$  is a scale parameter. The probability density function (pdf) is:

$$f(x) = \alpha^{-1} [1 - k(x - \zeta)/\alpha]^{(1/k)-1} [F(x)]^{1-h}. \tag{2}$$

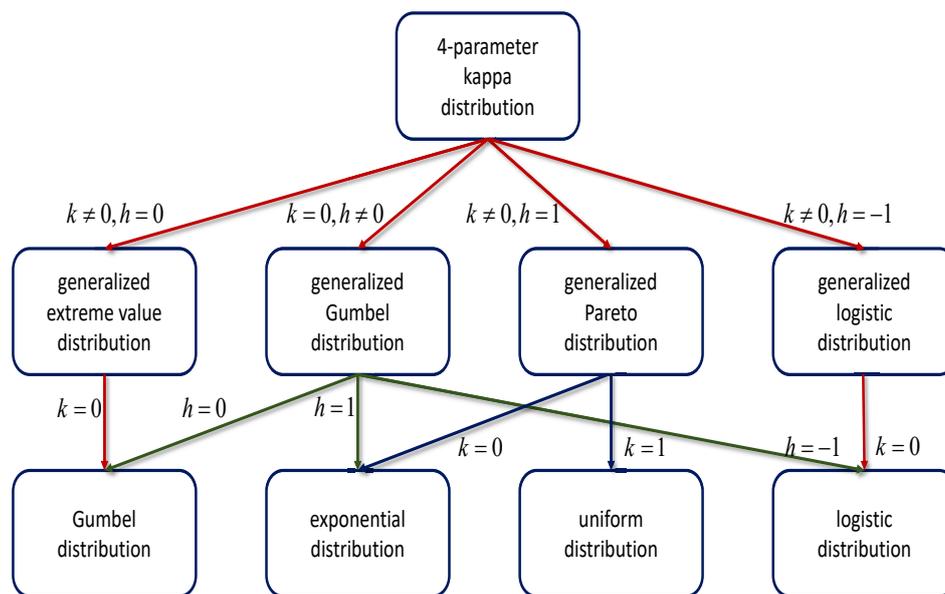
The changes in extremes are often described in terms of the changes in extreme quantiles. These values are obtained by solving the equation  $F(x) = 1 - 1/T$  with respect to  $x$  for a given year  $T$ ; that is, by inverting the following quantile function:

$$x(F) = \zeta + \frac{\alpha}{k} \left[ 1 - \left( \frac{1 - F(x)^h}{h} \right)^k \right]. \tag{3}$$

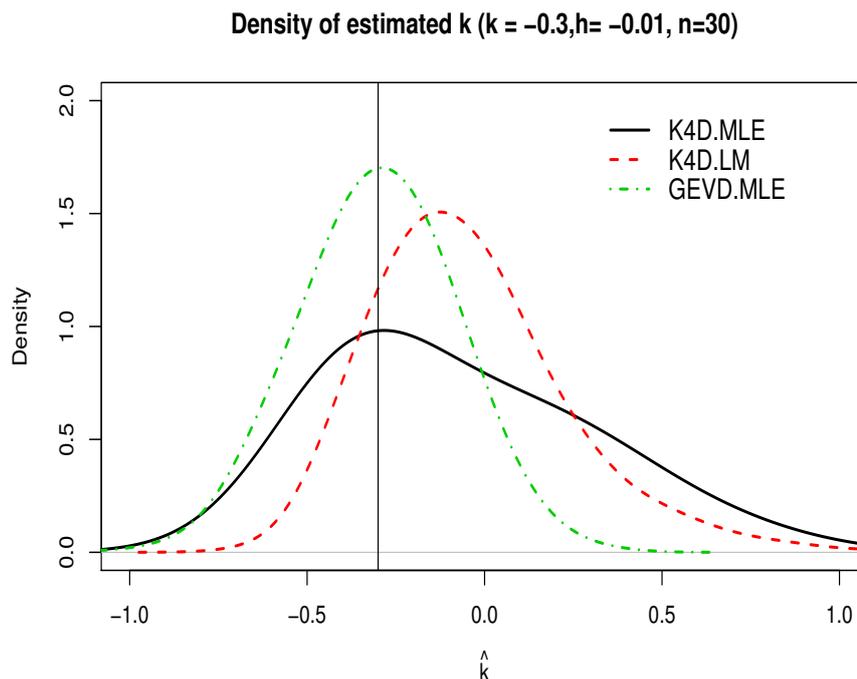
Here,  $T$  is called a return period. The value  $x(1 - 1/T)$  is referred to as the return level (or return value) associated with the return period  $T$ , because the value  $x(1 - 1/T)$  is expected to be exceeded on average once every  $T$  years [12]. For example, a 20-year return value is computed as the 95th percentile of the fitted K4D from  $x(0.95)$  (because  $1 - 1/20 = 0.95$ ) and a 50-year return level as the 98th percentile.

As shown in Figure 1, the K4D includes many distributions as special cases: the generalized Pareto distribution for  $h = 1$ , the GEVD [12,21] for  $h = 0$ , the generalized logistic distribution (GLD) for  $h = -1$ , and the generalized Gumbel distribution [22] for  $k = 0$ . Thus, the K4D is very flexible and widely applicable to many situations including not only extreme values, but also skewed data. Figure 2 shows density plots in which the parameters of the K4D are estimated by different methods. Figure S1 in the SM presents some shapes of the pdf of K4D for various combinations of two shape parameters ( $k, h$ ) while the scale and location parameters are fixed at  $\alpha = 1$  and  $\xi = 0$ .

Parida [3] and Park and Jung [4] employed this distribution to analyze annual maximum daily precipitation in India and in Korea, respectively. It has also been used for the analysis of extreme flood flow series [7] and maximum wind speed [9]. Hosking and Wallis [2], Wallis et al. [6], and Kjeldsen et al. [8] adopted the K4D for regional frequency analysis of daily rainfall extremes and flood flow in the U.K. Shin et al. [23] considered the r-largest K4D. Dupuis and Winchester [11] examined this distribution within the ranges of  $k \in (-0.5, 0.5)$  and  $h \in (-1.2, 1.2)$ . The range of  $k \in (-0.5, 0.5)$  is based on the fact that, when  $h = 0$ , the distribution has finite variance for  $k > -0.5$  and is regular (has finite Fisher information) for  $k < 0.5$  [12]. We therefore confine this study to within these ranges of parameters.



**Figure 1.** Relationship of the four-parameter kappa distribution to other three-parameter and two-parameter distributions with varying shape parameters  $k$  and  $h$ .



**Figure 2.** Probability density plots obtained by the kernel method in which the parameters are estimated using different methods (L-moment estimation, maximum likelihood estimation) from random samples of the four-parameter kappa distribution (K4D) with  $k = -0.3$  and  $h = -0.01$  and sample size  $n = 30$ . The vertical line represents the true value of  $k$ . GEVD: generalized extreme value distribution, LM: L-moment estimation, MLE: maximum likelihood estimation.

### 3. Maximum Likelihood and L-Moments Estimations

#### 3.1. Maximum Likelihood Estimation

The MLE is the most popular method of estimation and inference for parametric models. Suppose  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  is a sample of size  $n$  obtained from the K4D with pdf  $f(x)$ ; assuming  $k \neq 0$  and  $h \neq 0$ , the negative log-likelihood function of the parameters  $\theta = (\zeta, \alpha, k, h)$  is written as:

$$\ell(\theta|\mathbf{x}) = n \log \alpha - (1 - h) \sum_{i=1}^n \log F(x_i) - \frac{1 - k}{k} \sum_{i=1}^n \log G_i, \tag{4}$$

where  $G_i = 1 - k(x_i - \zeta)/\alpha$ . The constraints [1] for the MLE are that  $G_i$  is positive for all  $x_i$ , and  $x_i \leq U$  if  $k > 0$ ,  $x_i \geq E$  if  $h > 0$ , and  $x_i \geq U$  if  $h \leq 0, k < 0$ , where  $U = \zeta + \alpha/k, E = \zeta + \alpha(1 - h^{-h})/k$ .

The MLEs are obtained by minimizing Equation (4), setting the first partial derivatives of  $\ell(\theta|\mathbf{x})$  to zero with respect to  $\theta$ . Since no explicit minimizer is possible, a quasi-Newton algorithm is implemented using the “optim” function in the R program [24] to minimize Equation (4). Park and Park [25] calculated the MLE of the K4D parameters employing a penalty method.

We find that the MLE of the K4D with a small sample performs poorly and is unstable, especially when the scale parameter  $k$  is negative. It produces relatively large variances of estimated parameters compared to the LME (Figure 2). This poor performance of the MLE with a small sample may be due to an intrinsic property of the K4D or due to adding one more parameter from GEVD with more constraints, which makes it harder for the optimization routine to reach the true solution. The K4D however may result in less biased estimates than the GEVD does, by the bias-variance trade-off principle.

### 3.2. L-Moments Estimation

The main benefits of using LME are that the parameter estimates are more credible than the method of moments estimates, particularly for small samples, and are computationally more efficient than the MLE. LME is known to work better than the MLE for small sample sizes. Moreover, it is less sensitive to outliers [2].

Hosking [26] introduced L-moments as a linear combination of expectations of order statistics. The sample L-moment as an unbiased estimator of the population L-moment is calculated from a sample. The LME is acquired by solving a system of equations that equate population L-moments to the corresponding sample L-moments. The system of equations is sometimes solved numerically using Newton-Raphson iterations. However, the iterations sometimes do not converge to the solution. Dupuis and Winchester [11] reported that this non-feasibility can be observed in up to 40% of trials in the K4D. LME has been used widely in many fields, including environmental sciences [8,27]. The details are omitted here because it is a standard method for analyzing skewed data and extreme values. See [1,2] for the L-moments formula and the formula for obtaining the LME of the K4D.

The LME of the K4D was used for regional flood frequency analysis in Kjeldsen et al. [8] and Hosking and Wallis [2]. Parida [3] and Moses and Parida [9] used the LME of the K4D for modeling Indian summer monsoon rainfall and Botswana’s monthly maximum wind speed, respectively. Murshed et al. [7] considered the higher order L-moment estimator in the K4D with application to annual maximum flood and sea-level data. Dupuis and Winchester [11] compared the performance of the MLE and LME of the K4D. The disadvantage of LME is that it is difficult to apply it to non-stationary or multivariate data. For this study, we used the R package “lmom” developed by Hosking [26] to calculate the LME of the K4D.

## 4. Bayesian Inference

The main difference between classical frequentist theory and Bayesian estimation is that the latter considers parameters as random variables represented by a prior distribution. This prior distribution is incorporated with the likelihood to obtain the posterior of the parameter on which the statistical inference is based. We now set up a framework to incorporate prior information  $\pi(\theta)$  with the data. Let us assume that  $\mathbf{x} = (x_1, \dots, x_n)$  are independent realizations of a random variable whose probability density lives within a parametric family  $\{f(\mathbf{x}; \theta) : \theta \in \Theta\}$ .

The posterior distribution is proportional to the product of the likelihood function and the prior, but the computation of the normalizing constant  $\int_{\Theta} \pi(\theta)L(\theta|\mathbf{x})d\theta$  can be problematic, especially when  $\theta$  is a high-dimensional vector. Although this can be addressed in various ways, recent advances in the Markov chain Monte Carlo (MCMC) method have revolutionized the subject. It enables the statistical inference of models that would otherwise not be available [28]. Our approach of obtaining the posterior is the random walk Metropolis–Hastings algorithm. Specifically, Steps 3 and 4 described in the following subsection are for the conditional posterior distribution and for the incorporation with the parameter  $k$ .

### 4.1. Computation by Markov Chain Monte Carlo

Using the MCMC technique, we pursue generating (stationary) sequences of simulated values from the posterior distribution. The Metropolis–Hastings algorithm was developed by Metropolis et al. [29] and generalized by Hastings [30], and it is now the most popular MCMC method (see [31–33], for example).

The random variables  $\mathbf{x}$  are distributed as the K4D with the likelihood function  $L(k, h, \alpha, \zeta|\mathbf{x})$ . To specify the prior, the parametrization  $\phi = \log(\alpha)$  is applied because  $\alpha$  should be positive. The independent prior density is set as:

$$\pi(k, h, \phi, \zeta) = \pi_k(k)\pi_h(h)\pi_\phi(\phi)\pi_\zeta(\zeta), \tag{5}$$

where  $\pi_k(k)$ ,  $\pi_h(h)$ ,  $\pi_\phi(\phi)$ , and  $\pi_{\xi}(\xi)$  are the marginal priors of  $k$ ,  $h$ ,  $\phi$ , and  $\xi$ , respectively.

Our interest is the posterior density:

$$\pi(k, h, \phi, \xi | \mathbf{x}) \propto \pi(k, h, \phi, \xi) L(k, h, \phi, \xi | \mathbf{x}). \tag{6}$$

By Equation (5), the posterior is of the form:

$$\pi(k, h, \phi, \xi) L(k, h, \phi, \xi | \mathbf{x}) = \pi_k(k) \pi_h(h) \pi_\phi(\phi) \pi_{\xi}(\xi) L(k, h, \phi, \xi | \mathbf{x}). \tag{7}$$

Hence, the conditionals of the posterior distribution of parameters  $k$ ,  $h$ ,  $\phi$ , and  $\xi$  are as follows:

$$\pi(k | h, \phi, \xi) = \pi_k(k) L(k, h, \phi, \xi | \mathbf{x}), \tag{8}$$

$$\pi(h | k, \phi, \xi) = \pi_h(h) L(k, h, \phi, \xi | \mathbf{x}), \tag{9}$$

$$\pi(\phi | k, h, \xi) = \pi_\phi(\phi) L(k, h, \phi, \xi | \mathbf{x}), \tag{10}$$

$$\pi(\xi | k, h, \phi) = \pi_{\xi}(\xi) L(k, h, \phi, \xi | \mathbf{x}). \tag{11}$$

To simulate each of the above conditionals of the posterior distribution, a Metropolis step with a random walk update is employed, because the above form is a multivariate density [34].

In the MCMC algorithm, the value of  $r = 2000$  is the burn-in period, and the first  $r$  values are discarded from the chain. The maximum a posteriori (MAP) estimator is:

$$\hat{\theta} = \arg \max_{\theta} \pi(\theta^{(i)} | \mathbf{x}), \tag{12}$$

where  $i = r + 1, \dots, R$ , and  $R$  is the number of iterations in the Metropolis–Hastings algorithm. We set  $R = 10,000$ .

The full details of the random walk Metropolis–Hastings algorithm employed in this study are described as follows:

1. Start with an initial value  $k^{(0)}, h^{(0)}, \phi^{(0)}, \xi^{(0)}$ .
2. For  $i = 1, 2, \dots, R$
3. Generate a so-called “proposed value”  $k^* \sim N(k^{(i-1)}, s_k^2)$ , where the “proposed” standard deviation of  $k$ ,  $s_k$  is specified. Set:

$$r_k = \min \left\{ 1, \frac{\pi(k^* | h^{(i-1)}, \phi^{(i-1)}, \xi^{(i-1)})}{\pi(k^{(i-1)} | h^{(i-1)}, \phi^{(i-1)}, \xi^{(i-1)})} \right\}.$$

Accept:

$$k^{(i)} = \begin{cases} k^* & \text{with probability } r_k \\ k^{(i-1)} & \text{with probability } 1 - r_k. \end{cases}$$

4. Generate a proposed value  $h^* \sim N(h^{(i-1)}, s_h^2)$ , where the proposed standard deviation of  $h$ ,  $s_h$  is specified. Set:

$$r_h = \min \left\{ 1, \frac{\pi(h^* | k^{(i)}, \phi^{(i-1)}, \xi^{(i-1)})}{\pi(h^{(i-1)} | k^{(i)}, \phi^{(i-1)}, \xi^{(i-1)})} \right\}.$$

Accept:

$$h^{(i)} = \begin{cases} h^* & \text{with probability } r_h \\ h^{(i-1)} & \text{with probability } 1 - r_h. \end{cases}$$

5. Generate a proposed value  $\phi^* \sim N(\phi^{(i-1)}, s_\phi^2)$ , where the proposed standard deviation of  $\phi$ ,  $s_\phi$  is specified. Set:

$$r_\phi = \min \left\{ 1, \frac{\pi(\phi^* | k^{(i)}, h^{(i)}, \zeta^{(i-1)})}{\pi(\phi^{(i-1)} | k^{(i)}, h^{(i)}, \zeta^{(i-1)})} \right\}.$$

Accept:

$$\phi^{(i)} = \begin{cases} \phi^* & \text{with probability } r_\phi \\ \phi^{(i-1)} & \text{with probability } 1 - r_\phi. \end{cases}$$

6. Generate a proposed value  $\zeta^* \sim N(\zeta^{(i-1)}, s_\zeta^2)$ , where the proposed standard deviation of  $\zeta$ ,  $s_\zeta$  is specified. Set:

$$r_\zeta = \min \left\{ 1, \frac{\pi(\zeta^* | k^{(i)}, h^{(i)}, \phi^{(i)})}{\pi(\zeta^{(i-1)} | k^{(i)}, h^{(i)}, \phi^{(i)})} \right\}.$$

Accept:

$$\zeta^{(i)} = \begin{cases} \zeta^* & \text{with probability } r_\zeta \\ \zeta^{(i-1)} & \text{with probability } 1 - r_\zeta. \end{cases}$$

7. Increase the counter from  $i = i + 1$ , and repeat Step 2.
8. The final step is to transform our chain for  $\phi$  using  $\alpha^{(i)} = \exp(\phi^{(i)})$  to regain our original scale parameter.

Here, the “proposed” standard deviation ( $s_k$ ,  $s_h$ ,  $s_\phi$ , and  $s_\zeta$ ) was chosen by a calibration process to make the acceptance rates be not far from 30% such as between 25% and 35% [35]. Actually, at the very beginning of the algorithm, the initial values of standard deviations are given such that all are equal to 1.0. When the iterations of Metropolis–Hastings algorithm are done, the acceptance rate (AR) is provided. If the AR is greater than 35%, the standard deviations are updated to be larger than the previous ones. If the AR is smaller than 25%, those are updated to be smaller than the previous ones. Then Metropolis–Hastings algorithm is run again with the updated standard deviations, until the AR falls between 25% and 35%.

#### 4.2. Prior Specification

The principal argument against a Bayesian analysis is that the results are sometimes too sensitive to the choice of the prior distribution. Previous research on similar models or relevant data has provided some details on the prior distribution. Many researchers have applied the GEVD to hydrologic data. Some studies have offered a more exact constraint on the range of the shape parameters. Martins and Stedinger [36] found that  $-0.3$  to  $0$  is the most likely range for the shape parameter for GEVD. Sometimes, confusion arises regarding the sign of  $k$  between Martins and Stedinger [36] and Coles [12]. For consistency with the definition of the K4D, we follow the former.

Previous MLE research for long records of heavy rainfall in the U.K. have found the mean estimate of the shape parameter to be  $-0.1$  with a standard deviation of  $0.1$  [37]. Extensive empirical evidence was provided by Koutsoyiannis [38], who examined 169 datasets of extreme rainfall records worldwide of a duration of at least 100 years. The investigation concluded that the mean of the estimated shape parameter is about  $-0.15$ , with a small standard deviation. They also found that 91% of the MLEs for the shape parameter were negative, which adds further weight to the belief that the shape parameter is negative for heavy rainfall. Based on this information, we set the following priors:

- P1: Non-informative priors on all shape parameters:  $k \sim N(0, 1)$  and  $h \sim N(0, 2)$ . Given that the ranges of  $k$  and  $h$  are  $(-0.5, 0.5)$  and  $(-1.2, 1.2)$ , respectively, the variances of this prior make the distributions almost non-informative.
- P2: Priors P2 are based on the hydrologic experience that the shape parameters of the GEVD are between  $-0.3$  to  $0$  for rainfall data [36,38]:  $k \sim N(-0.15, 0.2)$  and  $h \sim N(-0.15, 0.5)$ .

- P3: Priors P3 are normal distributions with means of  $-0.15$  and  $0.15$ , respectively, but with different variances reflecting different levels of uncertainty because there is no previous evidence on  $h$ . Note that the sign for  $h$  is positive:  $k \sim N(-0.15, 0.01)$  and  $h \sim N(0.15, (0.25)^2)$ .
- P4: Tight normal distributions around the shape parameter to assume the best scenario:  $k \sim N(k_{opt}, 0.01)$  and  $h \sim N(h_{opt}, (0.25)^2)$ . In real applications, the mean values ( $k_{opt}$  and  $h_{opt}$ , respectively) of  $k$  and  $h$  are estimated by LME from data.
- P5: When  $k < 0$  and  $h < 0$ , negative priors correspond to  $k = -|\gamma_1|$  and  $h = -|\gamma_2|$ , where  $\gamma_1 \sim N(0, 1/3)$  and  $\gamma_2 \sim N(0, 2/3)$ .

The non-informative prior (P1) is least sensitive to the selection of priors and still covers the range of true shape parameters. We therefore expect the Bayes estimator with P1 to behave similarly to the MLE and LME, but it still supplies a more comprehensive summary of the analysis, such as posterior density estimates of parameters and return values. A feature of P4 is that the mean values of  $k$  and  $h$  are estimated by LME from data, which can be viewed as an empirical Bayes approach.

### 5. Simulation Study

We now compare the performances of the MLE, LME, and BEs for the K4D. Simulations are performed for sample sizes of 30, 50, and 100 generated from the K4D with different values of the shape parameters:  $k = \{-0.4, -0.2, 0.2, 0.4\}$  and  $h = \{-1.0, -0.6, -0.2, 0.2, 0.6, 0.9\}$ . The value  $h = 1.0$  is not included in this study because our pilot experiments of MCMC failed with a zero acceptance rate when  $h = 1.0$ . Thus, it was replaced by  $h = 0.9$ . The scale and location parameters are fixed at one and zero, respectively.

A total of 1000 samples are tried to calculate the bias (BIAS) and the root mean squared error (RMSE) of each parameter and the bias (BIASQ) and the root mean squared error (RMSEQ) of the 95th percentile estimators. The BIAS, RMSE, BIASQ, and RMSEQ are defined as:

$$\text{BIAS} = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i - \theta), \tag{13}$$

$$\text{RMSE} = \sqrt{\frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i - \theta)^2}, \tag{14}$$

$$\text{BIASQ} = \frac{1}{1000} \sum_{i=1}^{1000} (q_i^e - q^t), \tag{15}$$

$$\text{RMSEQ} = \sqrt{\frac{1}{1000} \sum_{i=1}^{1000} (q_i^e - q^t)^2}, \tag{16}$$

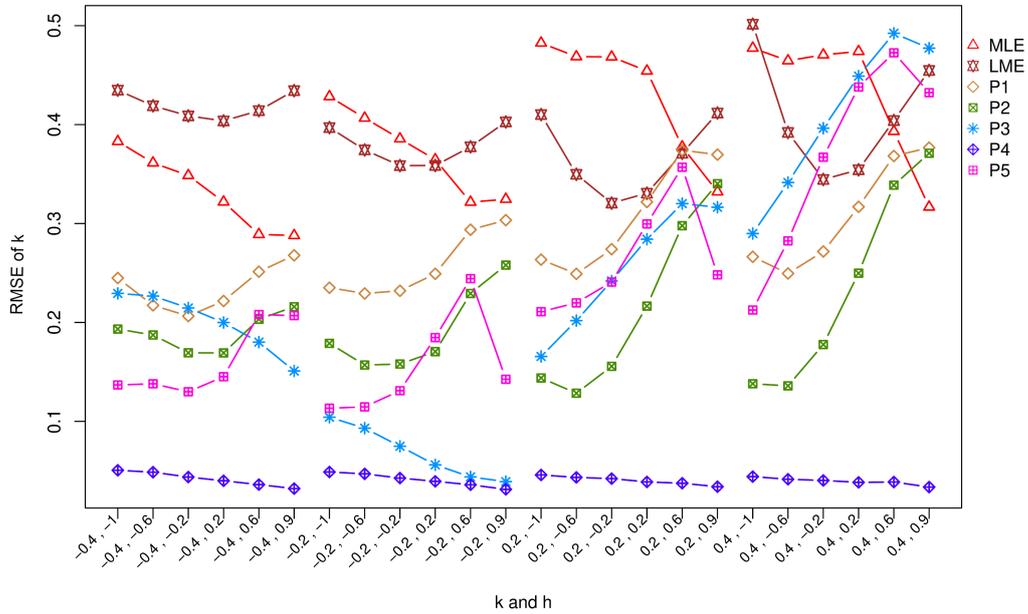
where  $\hat{\theta}$  and  $\theta$  are the estimated and true parameters, respectively, and  $q_i^e$  and  $q^t$  are the estimated and true quantiles, respectively.

Figures 3–6 show the results of the simulation for various combinations of  $k$  and  $h$  with sample size 30. The average CPU time for each method is shown in Table S1 in the SM (Supplementary Material) when the sample sizes of the simulation are 30, 50, 100, and 200.

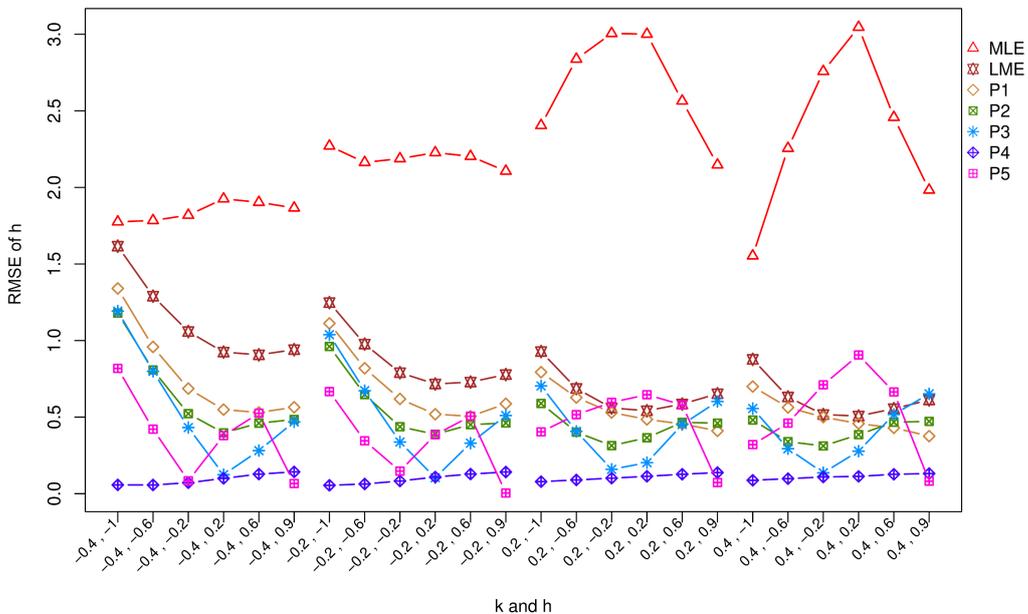
Figure S2 in the SM shows the estimation bias of the shape parameter of  $k$ . We can see that the MLE and LME are positively biased while the BEs are negatively biased. The BE with prior P4 yields consistently good estimates. All estimators have decreasing bias as  $k$  and  $h$  increase. It is interesting to note that the BE with P1 (non-informative prior) has a relatively small bias compared to the others. Figure S3 in the SM shows the estimation bias of  $h$ . The MLE is negatively biased, while LME is still positively biased. BEs have a similar pattern and relative performance as in Figure 4.

The comparisons of the RMSE for shape parameter  $k$  are given in Figure 3. For negative  $k$ , the MLE and LME work relatively poorly, but they do not work so bad for positive  $k$  compared to the BEs.

For positive  $k$ , the RMSE of the BEs increases as  $h$  increases. BEs work well in general except for high positive values of  $k$  and  $h$ . In Figure 4, the MLEs have the highest RMSE in estimating  $h$ . The BEs work better than LME for negative  $k$  and work similarly for LME for positive  $k$  in the RMSE sense. From the result based on the terms of the bias and RMSE for  $k$  and  $h$  estimations, we observe that the BEs work well.



**Figure 3.** The root mean squared error (RMSE) of the estimates of  $k$  under various combinations of  $k$  and  $h$  for sample size 30. The MLE, LME, and Pi are the maximum likelihood estimator, the method of L-moments estimator, and the Bayesian estimators with five priors Pi (P1 to P5), respectively.



**Figure 4.** Same as Figure 3, but for parameter  $h$ .

Figure 5 reveals the BIASQ, where the MLE, LME, and BEs with prior P4 have small biases, whereas the BEs with P1 and P3 work poorly. The comparison results in RMSEQ are presented in Figure 6. All the methods have decreasing RMSEQ as  $k$  and  $h$  increase. The MLE and BE with P1

do not work well for negative  $k$ , while the MLE and LME work well for positive  $k$ . The BE with P3 works well for negative  $k$ , while it works poorly for positive  $k$ . From the result based on the BIASQ and RMSEQ, we can infer that the MLE and LME work well in estimating quantiles when  $k$  is positive, even though these do not work well in estimating shape parameters in some cases. The prior P4 results in low RMSEQ values over all combinations of  $k$  and  $h$ . This is possibly because the mean values of  $k$  and  $h$  are estimated by LME from data. This result suggests that an empirical Bayesian method works well in the K4D.

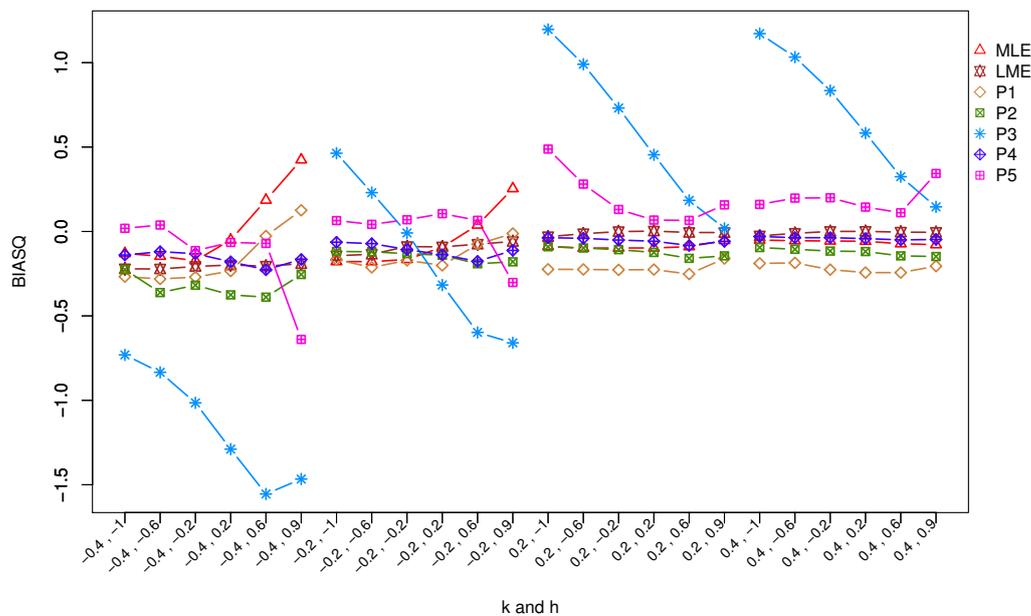


Figure 5. Same as Figure 3, but for the bias of the estimates for the 95th percentile (BIASQ).

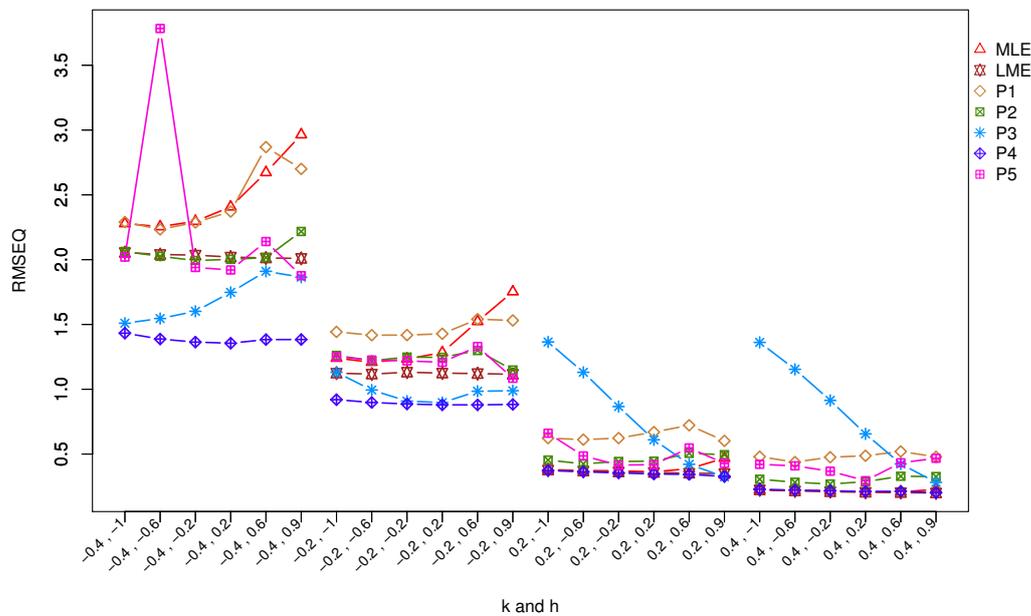


Figure 6. Same as Figure 3, but for the root mean squared error (RMSE) of the estimates of the 95th percentile (RMSEQ).

Overall, the BE with a non-informative prior (especially with P4) works well, even though there is no universal winner. The results of the Bayesian, maximum likelihood analysis, and L-moments

methods are consistent with each other, whereas the Bayesian analysis makes an easy transformation possible across scales, which is independent of the feasibility of LME and the asymptotic optimality of the MLE. Given that the MLE has a relatively larger variance and LME is sometimes non-feasible for the K4D and not applicable to non-stationary data, we can infer that the Bayesian framework (especially an empirical Bayes method) offers substantial benefits for the analysis of extreme events [12].

## 6. Applications to Real-World Data

### 6.1. Maximum Wind Speeds in Udon Thani, Thailand

Our example data consist of 26 annual maximum wind speeds (unit: km/h) in Udon Thani, Thailand, from 1990 to 2015 collected by the Meteorological Department of Thailand. The data are provided in the SM. Table 1 presents the summary statistics including parameter estimates with the standard error in parentheses, 20-year return levels, and 95% confidence intervals. Table 2 provides the goodness-of-fit measures including the Kolmogorov–Smirnov (K-S) test statistic, the Anderson–Darling (A-D) test statistic, and the information criteria such as the Akaike (AIC), Bayesian (BIC), and deviance information criterion (DIC). For all these measures, smaller is better. The MLE and LME for the GEV distribution are also included for comparison.

The DIC is a Bayesian analogy of the AIC. It has been successfully employed in model selection problems in which the posterior distributions of the models have been acquired by MCMC simulation. It is defined as:

$$DIC = D(\bar{\theta}) + 2p_D, \tag{17}$$

$$= \bar{D} + p_D, \tag{18}$$

where  $D(\bar{\theta}) = D(E_{\theta|x}[\theta])$  is the deviance at the mean posterior estimate,  $\bar{D} = E_{\theta|x}[D]$  is the expected value of the deviance over the posterior, and  $p_D = \bar{D} - D(\bar{\theta})$ . Here, the deviance is defined as  $D(\theta) = -2 \log L(\theta|x)$  [28,39].

A credible interval for the Bayesian estimate is obtained using the highest posterior density (HPD). The 95% confidence intervals (CI) for the 20-year return level of the MLE and LME are also obtained for comparison purposes. The profile likelihood method and a bias-corrected with acceleration constant bootstrap method are also considered in calculating the CI. The details of these methods are provided in the SM.

In Bayesian statistics, a credible (or posterior) interval is an interval within which an unobserved parameter value maintains a given probability [28,39]. Credible intervals are not unique on a posterior or a predictive distribution. Among some methods for defining an appropriate credible interval, the HPD interval chooses the narrowest interval, which consists of those values of highest probability density including the maximum a posteriori for a unimodal posterior distribution [39].

The HPD interval is detailed as follows: The 95% HPD consists of such values of  $\theta$  that have the minimal level of posterior credibility so that the total probability of all those  $\theta$  values is 0.95. A 100  $(1 - \beta)\%$  HPD region for  $\theta$  is a subset  $B \in \Theta$  defined by:

$$B = \{\theta \in \Theta : \pi(\theta|x) \geq \lambda(\beta)\}, \tag{19}$$

where  $\lambda(\beta)$  is the largest constant satisfying:

$$\int_B \pi(\theta|x) d\theta = 1 - \beta. \tag{20}$$

The value  $\lambda(\beta)$  is thought of as a horizontal line located over the posterior density, in which intersections with the posterior define regions with probability  $1 - \beta$  [39]. We calculated the HPD interval using the “coda” package in the R program, developed by Plummer et al. [40].

In Table 1, the 20-year return levels calculated from the BEs are greater than the corresponding values from MLE. This may be due to the implicitly permitted portion for uncertainty in parameter estimation [12]. In Table 2, the BEs with five priors yield  $p$ -values of K-S and of A-D larger than 0.7, which indicates that the maximum wind speed data are well fitted by the K4D with the BEs. Based on the length of the confidence intervals on the 20-year return levels, LME and the BE with P4 work well. The BEs with P1 and P4 have the smallest DIC values. In Tables 1 and 2, the values from the MLE, LME, AIC, and BIC are presented in this study only for the purpose of comparison with the BEs. In real Bayesian data analysis, analysts are usually not required to obtain these values, even though one can calculate and use them.

**Table 1.** Comparison of the estimates of the K4D and the standard errors in parentheses obtained by the MLE, L-moments, and Bayesian methods with the five priors (P1–P5) considered in the Section 4.2 and the confidence intervals (CI) of the 20-year return level (RL) for the annual maximum wind speed data (unit: km/h) at Udon Thani, Thailand, 1990–2015. The posterior mean and the credible (highest posterior density) interval are obtained for the Bayesian estimates.

| Methods  | Parameter Estimates |                  |                   |                   | CIs of 20-Year Return Level |                  |       |        |
|----------|---------------------|------------------|-------------------|-------------------|-----------------------------|------------------|-------|--------|
|          | $\hat{\xi}$         | $\hat{\alpha}$   | $\hat{k}$         | $\hat{h}$         | Lower                       | RL               | Upper | Length |
| Bayes P1 | 34.182<br>(0.020)   | 4.505<br>(0.020) | 0.195<br>(0.002)  | −0.323<br>(0.008) | 40.75                       | 43.95<br>(0.028) | 49.54 | 8.79   |
| Bayes P2 | 32.647<br>(0.016)   | 5.637<br>(0.017) | 0.248<br>(0.002)  | 0.009<br>(0.001)  | 41.07                       | 43.55<br>(0.028) | 50.47 | 9.40   |
| Bayes P3 | 32.930<br>(0.016)   | 4.993<br>(0.014) | −0.034<br>(0.001) | −0.116<br>(0.002) | 41.65                       | 47.11<br>(0.052) | 58.85 | 17.21  |
| Bayes P4 | 33.356<br>(0.017)   | 6.085<br>(0.014) | 0.414<br>(0.001)  | 0.106<br>(0.002)  | 41.30                       | 43.20<br>(0.016) | 46.86 | 5.57   |
| Bayes P5 | 34.324<br>(0.016)   | 3.488<br>(0.010) | −0.035<br>(0.001) | −0.612<br>(0.005) | 40.77                       | 45.63<br>(0.045) | 54.90 | 14.13  |
| MLE(K4D) | 32.617<br>(2.173)   | 6.688<br>(3.320) | 0.485<br>(0.268)  | 0.220<br>(0.381)  | 41.09                       | 43.15<br>(0.243) | 48.13 | 7.04   |
| LME(K4D) | 32.401<br>(1.096)   | 6.507<br>(5.126) | 0.401<br>(0.017)  | 0.219<br>(0.021)  | 41.69                       | 43.70<br>(0.043) | 46.42 | 4.73   |
| MLE(GEV) | 33.431<br>(1.148)   | 5.281<br>(0.841) | −0.363<br>(0.141) | -                 | 41.07                       | 43.02<br>(1.202) | 47.87 | 6.80   |
| LME(GEV) | 33.178<br>(0.012)   | 5.314<br>(0.009) | 0.287<br>(0.002)  | -                 | 42.72                       | 58.09<br>(0.131) | 97.24 | 54.52  |

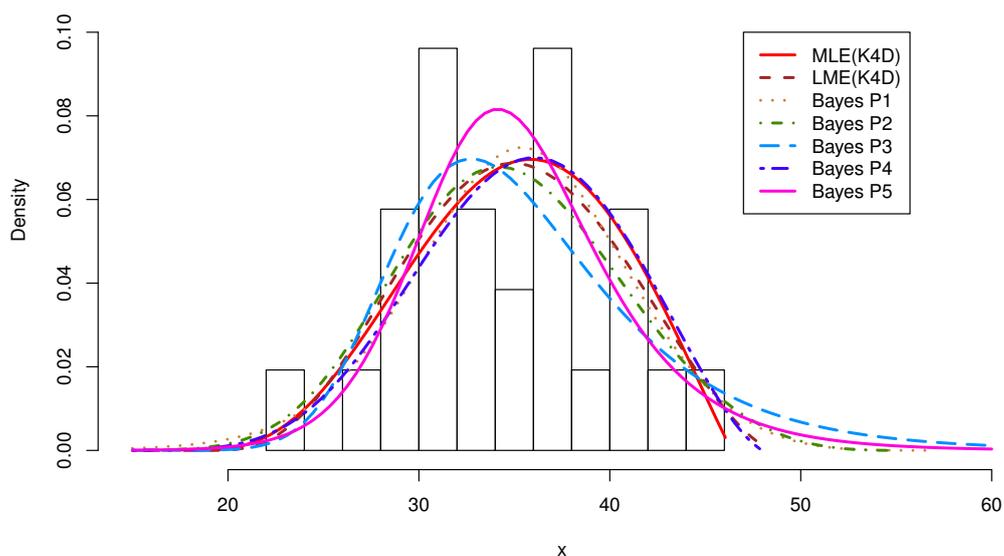
**Table 2.** Comparison of the K-S-test, A-D-test, AIC, BIC, and DIC obtained by MLE, L-moments, and Bayesian estimation methods with the five priors (P1–P5) considered in the Section 4.2 for the annual maximum wind speed data (unit: km/h) at Udon Thani, Thailand, 1990–2015. K-S: Kolmogorov-Smirnov, A-D: Anderson-Darling, AIC: Akaike information criterion, BIC: Bayesian information criterion, DIC: deviance information criterion, P:  $p$ -value.

| Methods  | K-S-Test |       | A-D-test |       | Criterion |        |        |
|----------|----------|-------|----------|-------|-----------|--------|--------|
|          | stat     | $p$   | stat     | $p$   | AIC       | BIC    | DIC    |
| Bayes P1 | 0.122    | 0.836 | 0.357    | 0.889 | -         | -      | 164.94 |
| Bayes P2 | 0.135    | 0.726 | 0.393    | 0.855 | -         | -      | 164.38 |
| Bayes P3 | 0.120    | 0.847 | 0.509    | 0.736 | -         | -      | 168.02 |
| Bayes P4 | 0.136    | 0.718 | 0.394    | 0.853 | -         | -      | 162.98 |
| Bayes P5 | 0.148    | 0.618 | 0.383    | 0.864 | -         | -      | 166.61 |
| MLE(K4D) | 0.114    | 0.886 | 0.331    | 0.912 | 165.80    | 170.83 | -      |
| LME(K4D) | 0.119    | 0.852 | 0.317    | 0.924 | 166.31    | 171.34 | -      |
| MLE(GEV) | 0.125    | 0.811 | 0.350    | 0.895 | 166.06    | 171.09 | -      |
| LME(GEV) | 0.177    | 0.389 | 1.388    | 0.206 | 189.33    | 194.36 | -      |

Figure 7 shows the probability density plots of the K4D fitted to the example data by various methods. Except the BEs with P3 and P5, other methods work similarly. Figure 8 shows the posterior densities and the HPD intervals of the 20-year return level for the annual maximum wind speeds in Udon Thani, obtained from the BEs. We have the narrowest and second narrowest HPD intervals from the BEs with P4 and P1, respectively. The BEs with P3 and P5 provide the widest HPD intervals. As observed in the simulation study, the BE with P4 (empirical Bayes method) works well in this example as well. For comparison purposes only, Figure S4 in the SM shows the profile log-likelihood function and a 95% confidence interval for the 20-year return value of the the annual maximum wind speeds in Udon Thani. The interval is [41.09, 48.13].

Figure 9 shows the densities of five priors (dashed lines), P1 to P5 considered in the Section 4.2, and of the posteriors (solid lines) for each parameter. Priors P3 and P4 for parameters  $h$  and  $k$  show that the posterior densities are strongly influenced by the assumed priors. These are not ideal, but inevitable for prior P4 because the mean values of  $h$  and  $k$  in P4 were estimated from data by the LME. The reason for prior P3 seems because the assumed variances of  $h$  and  $k$  in P3 were relatively small compared to those in priors P1, P2, and P5.

Using the method provided in Coles [12] (Section 9.1.2), the plots of the predictive distribution of a future annual maximum wind speed obtained by the BEs with five priors are drawn in Figure 10 on a  $\log_{10}$  return period scale. The predictive return levels of 1.6-year and 2.4-year return period for fives priors are between 40 to 55 km/h and 42 to 60 km/h, respectively, for the annual maximum wind speeds at Udon Thani.



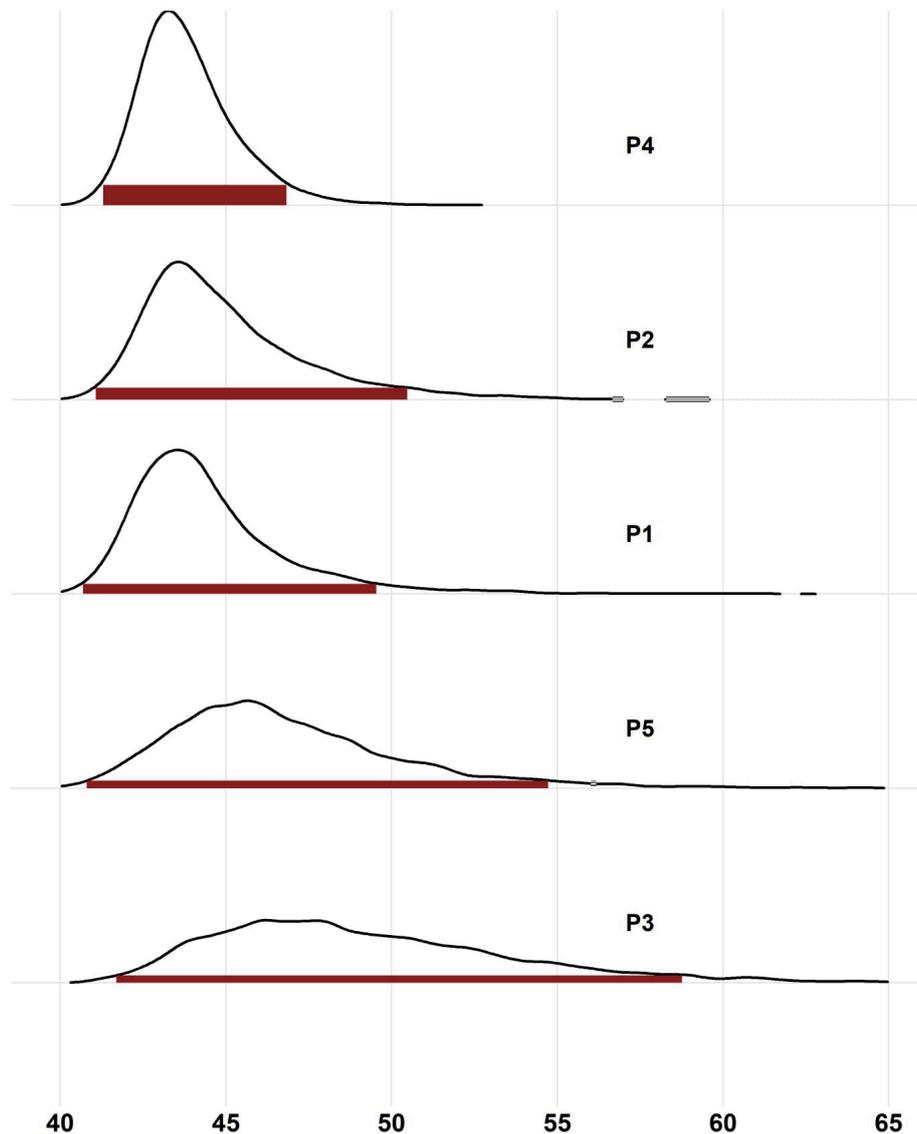
**Figure 7.** Probability density plots of four-parameter kappa distribution fitted to the annual maximum wind speeds at Udon Thani, Thailand, by various estimation methods.

The convergence of the MCMC chains can be assessed by some diagnostic measures including the potential scale reduction factor (PSRF) [41]. The estimated PSRF close to 1.0 indicates a successful convergence of the Markov chains. Table 3 provides the point estimates of PSRF and the upper bounds of its 95% confidence intervals by iterations in the Markov chains for each parameter based on the prior P1 considered in the Section 4.2 for the Udon Thani data. As iterations in the chain increase, the PSRF values decrease to 1.0. Figure 11 shows the trace plots of the estimated PSRF (solid line) and the upper bounds (dashed line) of its 95% confidence intervals in the Markov chains for each parameter. It appears to begin to be stationary after 500 iterations. Then, a burn-in is sufficient for 2000 iterations.

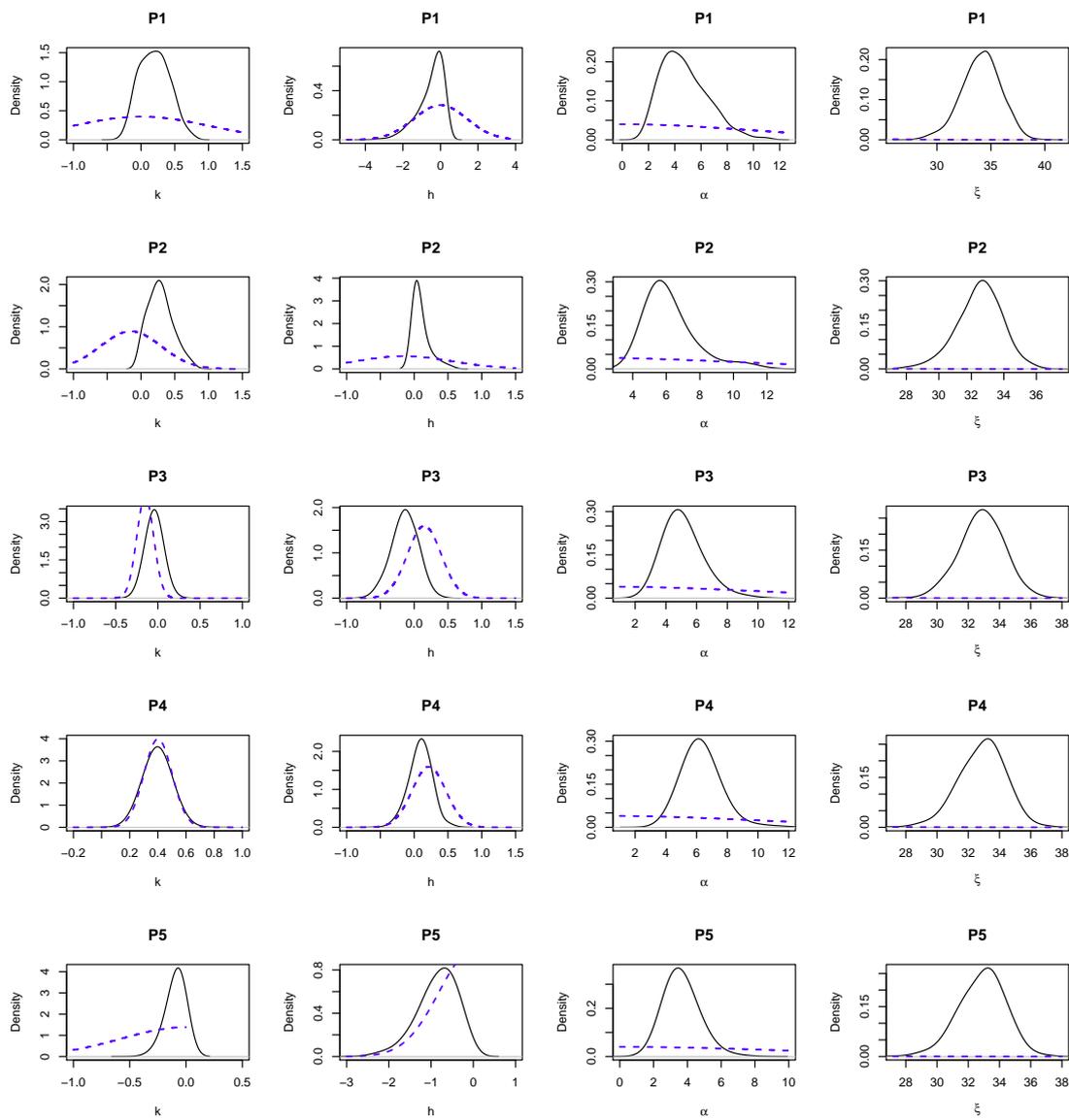
The effective numbers of the sample size (or independent draws) for each parameters  $k$ ,  $h$ ,  $\alpha$ , and  $\zeta$  are obtained as 200, 154, 105, and 255, respectively. These numbers are calculated by measuring

the autocorrelations within each sequence, based on the formula in [41]. If the effective number of independent draws is small, variances within the sequence will have a high sampling variability. Thus,  $\alpha$  among four parameters has the highest sampling variability. Moreover, the values generated by 10,000 iterations of the chain for diagnostic analysis are plotted in Figure S5, which also shows that the chain converges.

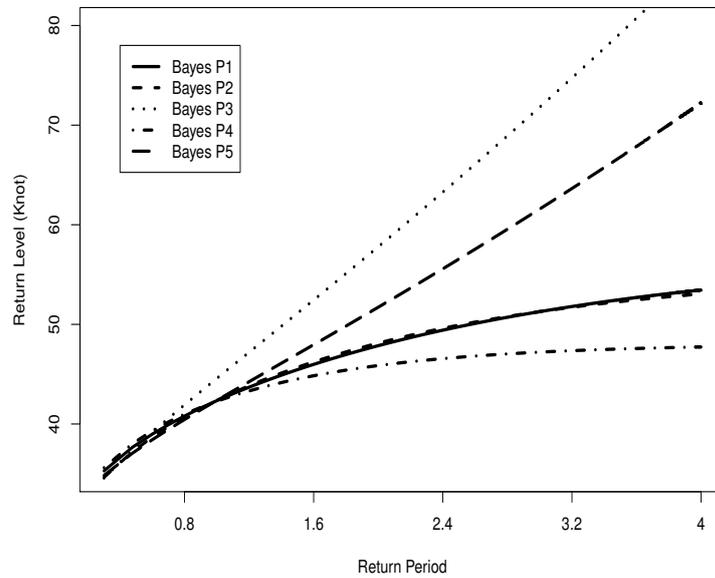
Figure S6 shows plots of the posterior correlations of the parameters. It seems there are weak non-linear correlations among parameters, except between  $h$  and  $\alpha$ . This may cause a problem resulting in long auto-correlation times. To make up for this correlation, it is practical to run MCMC sufficiently longer than the effective sample size [41].



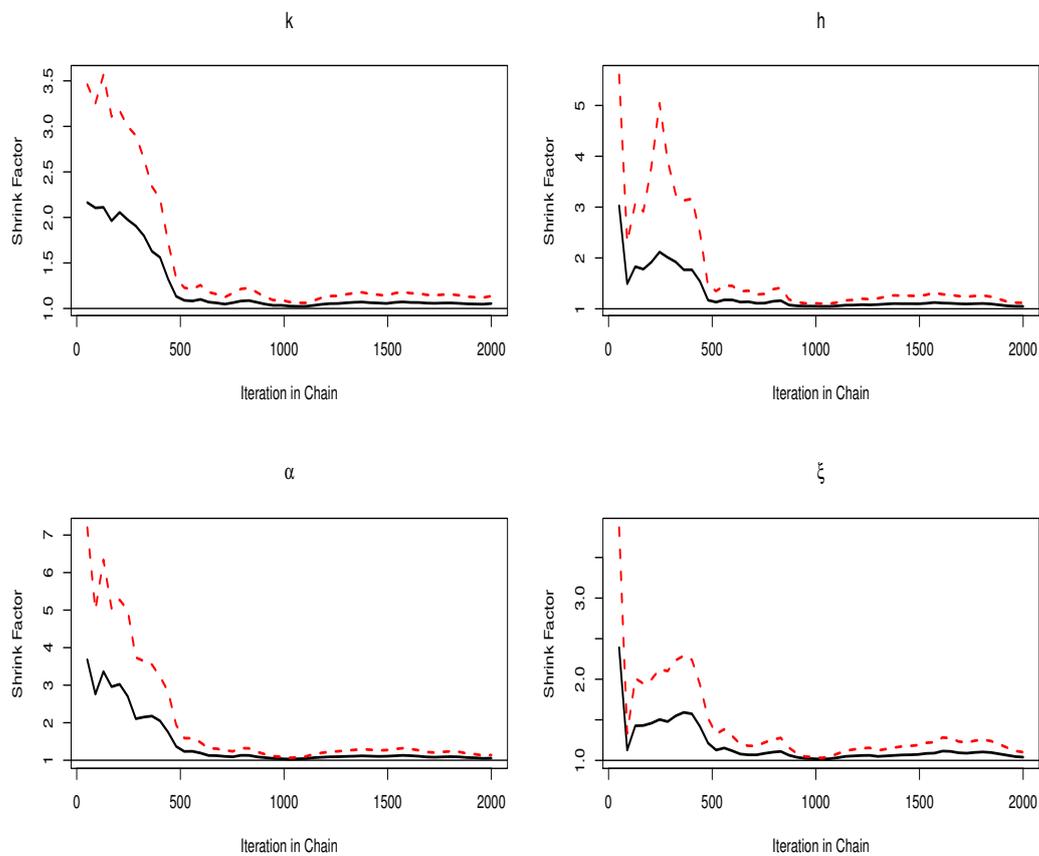
**Figure 8.** Posterior densities and highest posterior density intervals of the 20-year return levels (unit: km/h) obtained using the Bayesian estimation method with five priors (P1–P5) considered in the Section 4.2 for the annual maximum wind speed data at Udon Thani, Thailand.



**Figure 9.** Prior (dashed lines) and posterior (solid lines) densities for each of four parameters ( $k$ ,  $h$ ,  $\alpha$ , and  $\xi$ ) with the five priors (P1 to P5) considered in the Section 4.2 for the annual maximum wind speed data at Udon Thani, Thailand.



**Figure 10.** Predictive return levels against a  $\log_{10}$  return period scale in a Bayesian analysis with the five priors (P1 to P5) considered in the Section 4.2 for the annual maximum wind speeds at Udon Thani, Thailand.



**Figure 11.** The trace plots of the estimated potential scale reduction factors (solid line) and the upper bounds (dashed line) of its 95% confidence intervals by iterations in the Markov chains for each parameter from the annual maximum wind speed data at Udon Thani, Thailand.

**Table 3.** The point estimates of the potential scale reduction factors (PSRF) and the upper bounds of its 95% confidence intervals by iterations in the Markov chains for each parameter based on the prior P1 considered in the Section 4.2 for the annual maximum wind speed data at Udon Thani, Thailand.

| Iterations in Chain | Parameter   |             |             |             |
|---------------------|-------------|-------------|-------------|-------------|
|                     | $k$         | $h$         | $\alpha$    | $\zeta$     |
| 100                 | 1.62 (1.78) | 1.82 (2.04) | 2.07 (2.34) | 1.43 (1.55) |
| 200                 | 1.41 (1.57) | 1.57 (1.82) | 1.75 (2.02) | 1.35 (1.48) |
| 500                 | 1.11 (1.19) | 1.14 (1.24) | 1.17 (1.30) | 1.10 (1.17) |
| 1000                | 1.09 (1.18) | 1.09 (1.20) | 1.11 (1.22) | 1.07 (1.14) |
| 2000                | 1.05 (1.14) | 1.05 (1.12) | 1.06 (1.14) | 1.04 (1.10) |
| 5000                | 1.00 (1.01) | 1.00 (1.00) | 1.00 (1.00) | 1.02 (1.05) |

6.2. Maximum Sea-Levels at Fremantle, Australia: A Non-Stationary K4d

The next example data are the annual maximum sea-levels (unit: meter) recorded at Fremantle, near Perth, Western Australia, over the period 1897–1989. The data were analyzed in Coles [12] and are available from the website therein. As seen in Figure 12, there is a discernible increase in the data over time, although the increase seems marginal in recent years. A linear trend model of the location parameter is therefore introduced, with the scale and shape parameters remain constant:

$$\zeta(t) = \zeta_0 + \zeta_1 t. \tag{21}$$

Inference is now sought for the five parameters through the posterior distribution. Prior distributions are specified as follows, with a reparameterization  $\phi = \log(\alpha)$ :

$$\zeta_0 \sim \text{LogN}(0, 100), \tag{22}$$

$$\zeta_1 \sim \text{N}(0, 100), \tag{23}$$

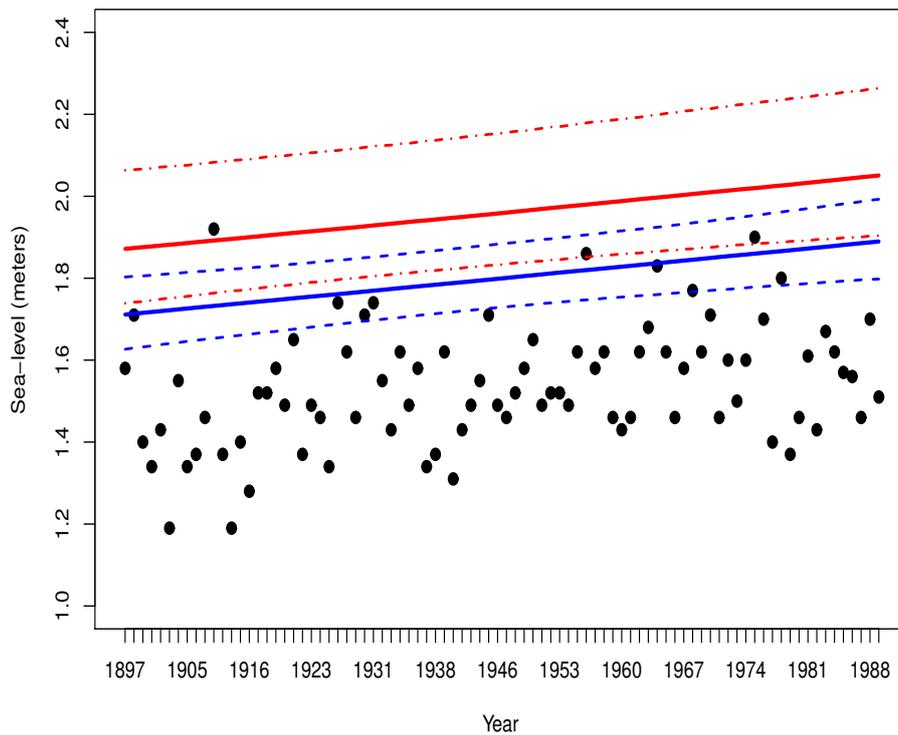
$$\phi \sim \text{N}(0, 100), \tag{24}$$

$$k \sim \text{N}(0, 1), \tag{25}$$

$$h \sim \text{N}(0, 2). \tag{26}$$

The distributions for  $\phi, k$ , and  $h$  are the same as the non-informative Prior 1 (P1) considered in the Section 4.2. To preserve the location parameter be positive, we set  $\zeta_0 \sim \text{LogN}(0, 100)$  instead of  $\text{N}(0, 100)$ .

Table 4 provides the estimates and DICs where the K4D and GEV distribution were fitted to the data. Based on the DIC, the Bayesian approach using the non-stationary K4D seems better than others. It is notable that a Bayes estimate of  $k$  on the non-stationary K4D is almost equal to zero, which results in the generalized Gumbel distribution [22]. Figure 13 shows the posterior pdf of each parameter and the 100-year return level. Given that the pdf of parameter  $\zeta_1$  is mostly concentrated on positive values, a positive trend may be suspected. Figure 12 shows the evolutions of the 20-year and 100-year return levels over time, with the 95% HPD intervals. The 20-year return level in 1960 is the level for 20 years corresponding to 1960. It is not calculated by extrapolating the time trend for the period 1960–1980, but calculated from the quantile function of the fitted non-stationary K4D. The above table and figures show that the non-stationary K4D with a Bayes estimate fits these sea-level data well.

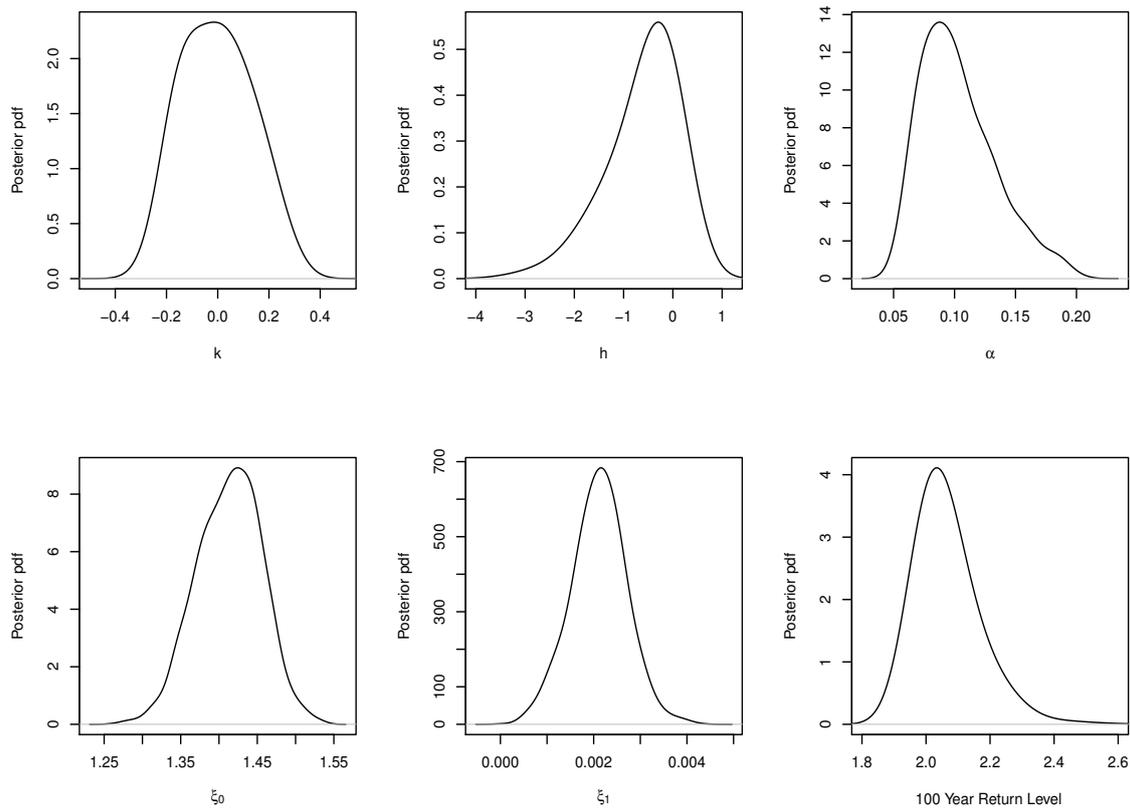


**Figure 12.** Evolutions of the 20-year and 100-year return levels in time, with 95% highest posterior density intervals (dashed lines) on the maximum sea-levels (unit: m) at Fremantle, Western Australia, 1897–1989. Blue lines for 20-year and red lines for 100-year return levels.

**Table 4.** Comparison of the estimates and deviance information criterion (DIC) obtained from the stationary and non-stationary GEVD and the K4D using Bayesian methods with the prior P1 given in the Section 4.2 for the annual maximum sea-levels (unit: meter) recorded at Fremantle, Western Australia, 1897–1989.

| Methods                     | Parameter Estimates |         |          |           |           | DIC       |
|-----------------------------|---------------------|---------|----------|-----------|-----------|-----------|
|                             | $k$                 | $h$     | $\alpha$ | $\zeta_0$ | $\zeta_1$ |           |
| GEV Bayes                   | −0.2296             | -       | 0.1424   | 1.4816    | -         | −81.1045  |
| GEV Bayes Non-stationary    | −0.1138             | -       | 0.1261   | 1.3639    | 0.0027    | −91.8363  |
| K4D Bayes P1                | 0.0101              | −0.2515 | 0.1115   | 1.5147    | -         | −84.8245  |
| K4D Bayes P1 Non-stationary | 0.0001              | −0.1926 | 0.0891   | 1.4219    | 0.0021    | −104.3389 |

Figure S7 in the SM shows the MCMC realizations of the posterior density function from 10,000 iterations for estimating the 100-year return level by the Bayesian method with the prior P1 given in the Section 4.2 using non-stationary K4D.



**Figure 13.** Posterior probability density function for each of five parameters ( $k$ ,  $h$ ,  $\alpha$ ,  $\zeta_0$ , and  $\zeta_1$ ) and the 100-year return level (lower rightest figure) obtained by the Bayesian method with the prior P1 given in the Section 4.2 using the non-stationary four-parameter kappa distribution, for the annual maximum sea-levels (unit: m) recorded at Fremantle, Western Australia, 1897–1989.

### 7. Conclusions and Discussion

This study considers Bayesian inference with five prior distributions for the K4D. A random walk Metropolis–Hastings algorithm is employed for simulating the posterior distribution. The properties of the BEs with five priors are verified by simulation and compared to those of the MLE and LME. The simulation results reveal that even though there is no universal winner among the considered estimators, the BE with prior P4 works well. This is because the mean values of  $k$  and  $h$  are estimated by LME from data, which suggests that an empirical Bayesian method works well in the K4D. The result of Bayesian estimation, maximum likelihood analysis, and the L-moments method are consistent with each other, whereas the Bayesian analysis makes an easy transformation possible across scales that does not depend on the feasibility of LME and the asymptotic optimality of the MLE.

In real-world examples, the annual maximum wind speed at Udon Thani, Thailand, was analyzed by the K4D. The data were well fitted by the K4D with the BEs. Based on the length of the confidence intervals on the 20-year return levels, LME and the BE with P4 work well. In another example with the annual maximum sea-level data in Fremantle, Australia, non-stationarity was modeled through a linear trend in time on the location parameter. These examples illustrate the improved efficiency of the Bayesian procedure in the parameter estimation of the K4D.

In general, the expected benefit of the Bayesian analysis is the improvements in the precision of the parameter estimates over the MLE and LME. One reason the BEs work better than the MLE and LME for a shape parameter of negative  $k$  is because the mean values of the priors of  $k$  are set to negative. These priors function well regarding the samples derived from the K4D with negative  $k$ . If we also take a prior that reflects a case of positive  $k$ , the result may be improved.

The value  $h = 1.0$  was not included in the simulation study because our pilot experiments of MCMC failed with a zero acceptance rate when  $h = 1.0$ . We were not successful in finding the reason why MCMC did not work in the case when  $h = 1$ . It may need an improved method as in [42]. We leave this task for future study.

In this study, the prior distribution of two shape parameters for the K4D were specified as having independent distributions. Future studies should find a bivariate prior for K4D that incorporates the joint information on  $k$  and  $h$ . The R code used to calculate the BEs of the K4D is available in the SM.

**Supplementary Materials:** The following supporting information is available as part of the online article: [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1). Figure S1. Shape of the probability density functions of the four-parameter kappa distribution with different combinations of  $k$  and  $h$  at  $\alpha = 1$  and  $\zeta = 0$ . Figure S2. Average of the estimate bias (BIAS) of  $k$  under various combinations of  $k$  and  $h$  for sample size 30. The acronyms MLE, LME, and Pi represent the maximum likelihood estimator, the method of L-moments estimator, and the Bayesian estimator with prior Pi, respectively. Figure S3. Same as Figure S2, but the BIAS for  $h$ . Figure S4. Profile likelihood and 95% confidence interval for the 20-year return levels of the K4D for the annual maximum wind speeds at Udon Thani, Thailand, where the parameters are estimated using the maximum likelihood method. Figure S5. Trace plots of the K4D parameters using five priors for the annual maximum wind speeds data at Udon Thani, Thailand. Figure S6. The posterior correlations of the parameters obtained from the annual maximum wind speed at Udon Thani, Thailand. Figure S7. MCMC realizations of posterior density function and 100-year return level with the nonstationary prior P1. Table S1. Average CPU time per case (seconds/data) using the Intel(R) Core(TM) i5-6300HQ CPU @ 2.30 GHz (RAM 8 GB). R program. R code to calculate the Bayesian estimates of the four-parameter kappa distribution.

**Author Contributions:** Conceptualization, P.S. and J.-S.P.; methodology, P.S., P.B., and J.-S.P.; software, P.S.; data curation, P.B.; writing, original draft preparation, P.S. and J.-S.P.; writing, review and editing, P.B. and J.-S.P.; visualization, P.S.; supervision, P.B.; communication, P.B.; funding acquisition, P.S., P.B., and J.-S.P. All authors read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Research Foundation (NRF) of Korea with Grant Number NRF-2016K2A9A1A09914124. Park's research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT No.2016R1A2B4014518, 2020R1I1A3069260) and by BK21 FOUR funded by the Ministry of Education, Korea (I20SS7609010). Piyapatr's work was supported by Mahasarakham University funded by the National Research Council of Thailand (NRCT).

**Acknowledgments:** The authors would like to thank the reviewers, the guest Editors for the Special Issue, and the Associate Editor for their helpful comments and suggestions, which greatly improved the presentation of this paper. This study began when the first author visited the Department of Statistics, Chonnam National University, for one year. He would like to thank the professors and staff, especially the graduate students Yonggwon Shin and Yire Shin, at DS/CNU, for their hospitality and support. We thank all contributors to the numerical R packages, which were crucial for this work. The authors are grateful to the Northeastern Meteorological Center (Upper Part) in Khon Kean province of Thailand for providing the wind speed data for Udon Thani.

**Conflicts of Interest:** The authors declare that they have no conflict of interest.

## References

- Hosking, J.R.M. The four-parameter kappa distribution. *IBM J. Res. Dev.* **1994**, *38*, 251–258. [[CrossRef](#)]
- Hosking, J.R.M.; Wallis, J.R. *Regional Frequency Analysis: An Approach Based on L-Moments*; Cambridge University Press: Cambridge, UK, 1997.
- Parida, B. Modelling of Indian summer monsoon rainfall using a four-parameter Kappa distribution. *Int. J. Climatol.* **1999**, *19*, 1389–1398. [[CrossRef](#)]
- Park, J.S.; Jung, H.S. Modelling Korean extreme rainfall using a Kappa distribution and maximum likelihood estimate. *Theor. Appl. Climatol.* **2002**, *72*, 55–64. [[CrossRef](#)]
- Singh, V.; Deng, Z. Entropy-based parameter estimation for kappa distribution. *J. Hydrol. Eng.* **2003**, *8*, 81–92. [[CrossRef](#)]
- Wallis, J.; Schaefer, M.; Barker, B.; Taylor, G. Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 415–442. [[CrossRef](#)]
- Murshed, S.; Seo, Y.A.; Park, J.S. LH-moment estimation of a four parameter kappa distribution with hydrologic applications. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 253–262. [[CrossRef](#)]
- Kjeldsen, T.R.; Ahn, H.; Prosdocimi, I. On the use of a four-parameter kappa distribution in regional frequency analysis. *Hydrol. Sci. J.* **2017**, *62*, 1354–1363. [[CrossRef](#)]

9. Moses, O.; Parida, B.P. Statistical Modelling of Botswana's Monthly Maximum Wind Speed using a Four-Parameter Kappa Distribution. *Am. J. Appl. Sci.* **2016**, *13*, 773–778. [[CrossRef](#)]
10. Blum, A.G.; Archfield, S.A.; Vogel, R.M. On the probability distribution of daily streamflow in the United States. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 3093–3103. [[CrossRef](#)]
11. Dupuis, D.; Winchester, C. More on the four-parameter kappa distribution. *J. Stat. Comput. Simul.* **2001**, *71*, 99–113. [[CrossRef](#)]
12. Coles, S. *An Introduction to Statistical Modeling of Extreme Values*; Springer: London, UK, 2001.
13. Coles, S.G.; Powell, E.A. Bayesian Methods in Extreme Value Modelling: A Review and New Developments. *Int. Stat. Rev.* **1996**, *64*, 119–136. [[CrossRef](#)]
14. Coles, S.G.; Tawn, J.A. A Bayesian Analysis of Extreme Rainfall Data. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1996**, *45*, 463–478. [[CrossRef](#)]
15. Coles, S.G.; Tawn, J.A. Bayesian modelling of extreme surges on the UK east coast. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* **2005**, *363*, 1387–1406. [[CrossRef](#)] [[PubMed](#)]
16. Vidal, I. A Bayesian analysis of the Gumbel distribution: An application to extreme rainfall data. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 571–582. [[CrossRef](#)]
17. Lima, C.H.; Kwon, H.H.; Kim, J.Y. A Bayesian beta distribution model for estimating rainfall IDF curves in a changing climate. *J. Hydrol.* **2016**, *540*, 744–756. [[CrossRef](#)]
18. Lima, C.H.; Lall, U.; Troy, T.; Devineni, N. A hierarchical Bayesian GEV model for improving local and regional flood quantile estimates. *J. Hydrol.* **2016**, *541*, 816–823. [[CrossRef](#)]
19. Fawcett, L.; Walshaw, D. Sea-surge and wind speed extremes: Optimal estimation strategies for planners and engineers. *Stoch. Environ. Res. Risk Assess.* **2016**, *30*, 463–480. [[CrossRef](#)]
20. Fawcett, L.; Green, A.C. Bayesian posterior predictive return levels for environmental extremes. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 2233–2252. [[CrossRef](#)]
21. Russell, B.T. Investigating precipitation extremes in South Carolina with focus on the state's October 2015 precipitation event. *J. Appl. Stat.* **2019**, *46*, 286–303. [[CrossRef](#)]
22. Jeong, B.Y.; Murshed, M.S.; Seo, Y.A.; Park, J.S. A three-parameter kappa distribution with hydrologic application: A generalized gumbel distribution. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 2063–2074. [[CrossRef](#)]
23. Shin, Y.; Busababodhin, P.; Park, J.S. The r-largest four parameter kappa distribution. *arXiv* **2020**, arXiv:2007.12031
24. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
25. Park, J.S.; Park, B.J. Maximum likelihood estimation of the four-parameter Kappa distribution using the penalty method. *Comput. Geosci.* **2002**, *28*, 65–68. [[CrossRef](#)]
26. Hosking, J.R.M. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order-Statistics. *J. R. Stat. Soc. Ser. B (Methodol.)* **1990**, *52*, 105–124. [[CrossRef](#)]
27. Šimková, T. Confidence intervals based on L-moments for quantiles of the GP and GEV distributions with application to market-opening asset prices data. *J. Appl. Stat.* **2020**, 1–28. [[CrossRef](#)]
28. Bernardo, J.M.; Smith, A.F.M. *Bayesian Theory*; John Wiley and Sons: Hoboken, NJ, USA, 1994. [[CrossRef](#)]
29. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [[CrossRef](#)]
30. Hastings, W.K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **1970**, *57*, 97–109. [[CrossRef](#)]
31. Saraiva, E.F.; Suzuki, A.K.; Milan, L.A. Bayesian Computational Methods for Sampling from the Posterior Distribution of a Bivariate Survival Model, Based on AMH Copula in the Presence of Right-Censored Data. *Entropy* **2018**, *20*, 642. [[CrossRef](#)] [[PubMed](#)]
32. Attia, A.; Moosavi, A.; Sandu, A. Cluster Sampling Filters for Non-Gaussian Data Assimilation. *Atmosphere* **2018**, *9*, 213. [[CrossRef](#)]
33. Ariza-Hernandez, F.J.; Arciga-Alejandre, M.P.; Sanchez-Ortiz, J.; Fleitas-Imbert, A. Bayesian Derivative Order Estimation for a Fractional Logistic Model. *Mathematics* **2020**, *8*, 109. [[CrossRef](#)]
34. Wasserman, L. *All of Statistics: A Concise Course in Statistical Inference*; Springer: Berlin, Germany, 2004.
35. Ntzoufras, I. *Bayesian Modeling Using WinBUGS*; Wiley: Hoboken, NJ, USA, 2009.

36. Martins, E.S.; Stedinger, J.R. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.* **2000**, *36*, 737–744. [[CrossRef](#)]
37. Wilson, P.S.; Toumi, R. A fundamental probability distribution for heavy rainfall. *Geophys. Res. Lett.* **2005**, *32*, L14812, [[CrossRef](#)]
38. Koutsoyiannis, D. Statistics of extremes and estimation of extreme rainfall: (ii) Empirical investigation of long rainfall records. *Hydrol. Sci. J.* **2004**, *49*, 591–610. [[CrossRef](#)]
39. Carlin, B.P.; Louis, T.A. *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed.; Texts in Statistical Science, Chapman and Hall: New York, NY, USA, 2000.
40. Plummer, M.; Best, N.; Cowles, K.; Vines, K. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* **2006**, *6*, 7–11.
41. Gelman, A.; Carlin, J.; Stern, H.; Dunson, D.; Vehtari, A.; Rubin, D. *Bayesian Data Analysis*, 3rd ed.; CRC Press: Boca Raton, FL, USA, 2013.
42. Liu, Y.; Lu, M.; Huo, X.; Hao, Y.; Gao, H.; Liu, Y.; Fan, Y.; Cui, Y.; Metivier, F. A Bayesian analysis of Generalized Pareto Distribution of runoff minima. *Hydrol. Process.* **2016**, *30*, 424–432. [[CrossRef](#)]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).