Supplementary Material: Bayesian Inference in Extremes using the Four-parameter Kappa Distribution

Palakorn Seenoi[a], Piyapatr Busababodhin[b],
Jeong-Soo Park[c]

a: Department of Statistics, Khon Kaen University, Khon Kaen, Thailand;

b: Department of Mathematics, Mahasarakham University, Maha Sarakham, Thailand;

c: Department of Mathematics and Statistics, Chonnam National University, Gwangju, Korea.

December 4, 2020

1 Criteria for model selection

Kolmogorov-Smirnov ($K-S$) test statistic, Anderson-Darling ($A-D$) test statistic, Bayesian information criterion, Akaike information criterion, and Deviance information criterion are applied to compare the goodness-of-fit of parameter estimates of Bayesian estimates with five priors, the MLE and the LME for the K4D. The estimation for GEVD is provided for comparison.

The formula for these criteria are as follows:

Kolmogorov-Smirnov test ($K-S$) is defined as [4],

$$K-S = \Big(\max_{1\leq i\leq n} \big[\max\{ \frac{i}{n} - F_0(X_{(i)}),\, F_0(X_{(i)}) - \frac{i-1}{n} \}\big] \Big)^2, \tag{1}$$

where $F_0(X_{(i)})$ is distribution function of $X_{(i)}$, $i = 1,2,...,n$.

Anderson-Darling test ($A-D$) is defined as [4],

$$A-D = -\frac{2}{n}\sum_{i=1}^{n} \big[\big(i-\frac{1}{2}\big) \log F_0(X_{(i)}) + \big(n-i+\frac{1}{2}\big) \log F_0(X_{(i)}) \big] - n, \tag{2}$$

where $F_0(X_{(i)})$ is distribution function of $X_{(i)}$, $i = 1,2,...,n$.l.

Akaike information criterion (AIC) is defined as [1],

$$AIC = 2p - 2\log L(\hat{\theta}|\boldsymbol{x}), \tag{3}$$

where $\hat{\theta}$ are the estimated parameter values and p is the number of the estimated parameters in the model.

Bayesian information criterion (BIC) is defined as [1],

$$BIC = \log(n)p - 2\log L(\hat{\theta}|\boldsymbol{x}), \tag{4}$$

where n is the number of data points in $\boldsymbol{x}$.

2 Return value and confidence interval

The 20-year return value (or level) and 95% confidence interval for 20-year return value of Bayesian, MLE and L-moments are obtained by the highest posterior density (HPD), profile likelihood and bias-corrected with acceleration constant bootstrap sampling methods, respectively.

The calculation of 95% confidence intervals for 20-year return value of the MLE is obtained from the profile likelihood method. The profile likelihood interval is detailed as follows [3]: The log-likelihood for θ is written as $\ell(\theta_i, \theta_{-i})$, in which θ_{-i} indicates all components of θ except θ_i. The profile log-likelihood for θ_i is written as

$$\ell_p(\theta_i) = \max_{\theta - i} \ell(\theta_i, \theta_{-i}). \tag{5}$$

We can obtain 95% confidence interval for 20-year return value, $x_{0.05}$, of the K4D model by reparameterization of quantile function in Eq. (6). Reparameterization is straightforward:

$$\xi = x_{0.05} - \frac{\alpha}{k}\left[1 - \left(\frac{1-(1-0.05)^h}{h}\right)^k\right], \tag{6}$$

so that $x_{0.05}$ is one of the model parameters [3]. Replacement of ξ in the log-likelihood function of the K4D with Eq. (6) having the desired effect of expressing the K4D model in terms of the reparameters ($x_{0.05}$, α, k, h).

To acquire the profile likelihood for 20-year return value $x_{0.05}$, we fix $x_{0.05}$ and maximize the log-

n	Methods		
	MLE	LME	Bayes
30	0.3282	0.0098	2.2646
50	0.4741	0.0098	3.4925
100	0.5797	0.0101	5.9212
200	0.8470	0.0114	11.8316

Table S1: Average of the Intel(R) Core(TM) i5-6300HQ CPU @ 2.30GHz (RAM 8 GB) CPU's time per case (Second/data)

likelihood with respect to the other parameters $(\alpha,\ k,\ h)$. The corresponding maximized values of the log-likelihood form the profile log-likelihood for $x_{0.05}$. This procedure is repeated for a range of values of $x_{0.05}$. Next, drawing the curve of the profile log-likelihood for $x_{0.05}$, the 95% confidence interval for 20-year return value is acquired, according to Coles [3], by drawing a horizontal line at height of $0.5 \times \chi_1^2(0.05)$ below the peak of the profile log-likelihood function, as in Figure S4, and reading off the points of intersection. Here, $\chi_1^2(0.05)$ is the 95 percentile of a χ_1^2 distribution.

The 95% confidence interval for 20-year return values of the L-moments is obtained by using bias-corrected with acceleration constant (BCa) bootstrap sampling method [2].

Figure S1 presents some shapes of the pdf of K4D for various combinations of two shape parameters $(k,\ h)$ where the scale and location parameters are fixed at $\alpha = 1$ and $\xi = 0$.

3 Maximum wind speeds data

Example data consists of 26 annual maximum wind speeds; there are 34, 41, 33, 31, 43, 40, 32, 29, 32, 38, 23, 38, 32, 45, 38, 33, 32, 30, 28, 41, 38, 41, 38, 36, 35 and 30 (unit: kilometres/hour), in Udon Thani, Thailand, from 1990 to 2015 which were collected by the Meteorological department of Thailand.

Figure S**??** shows plots of the posterior correlations of the parameters [**?**].

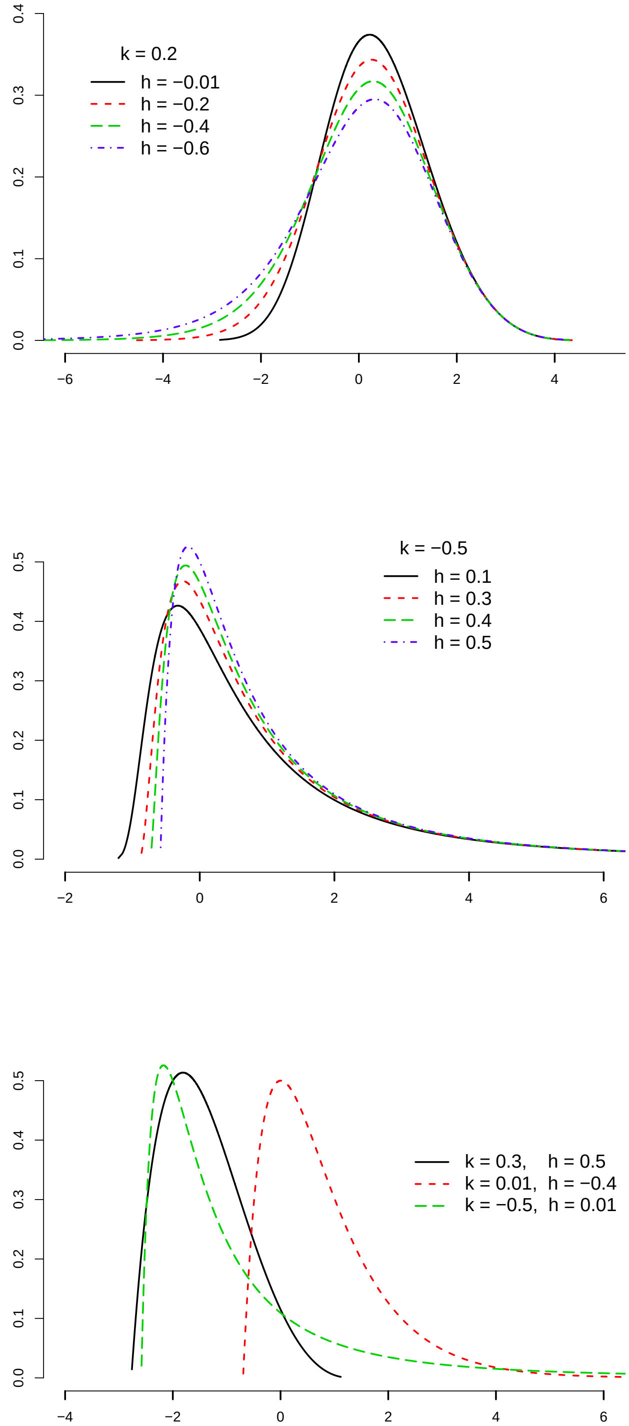

Figure S1: Shapes of the probability density functions of the four-parameter kappa distribution with different combinations of k and h at $\alpha = 1$ and $\xi = 0$.

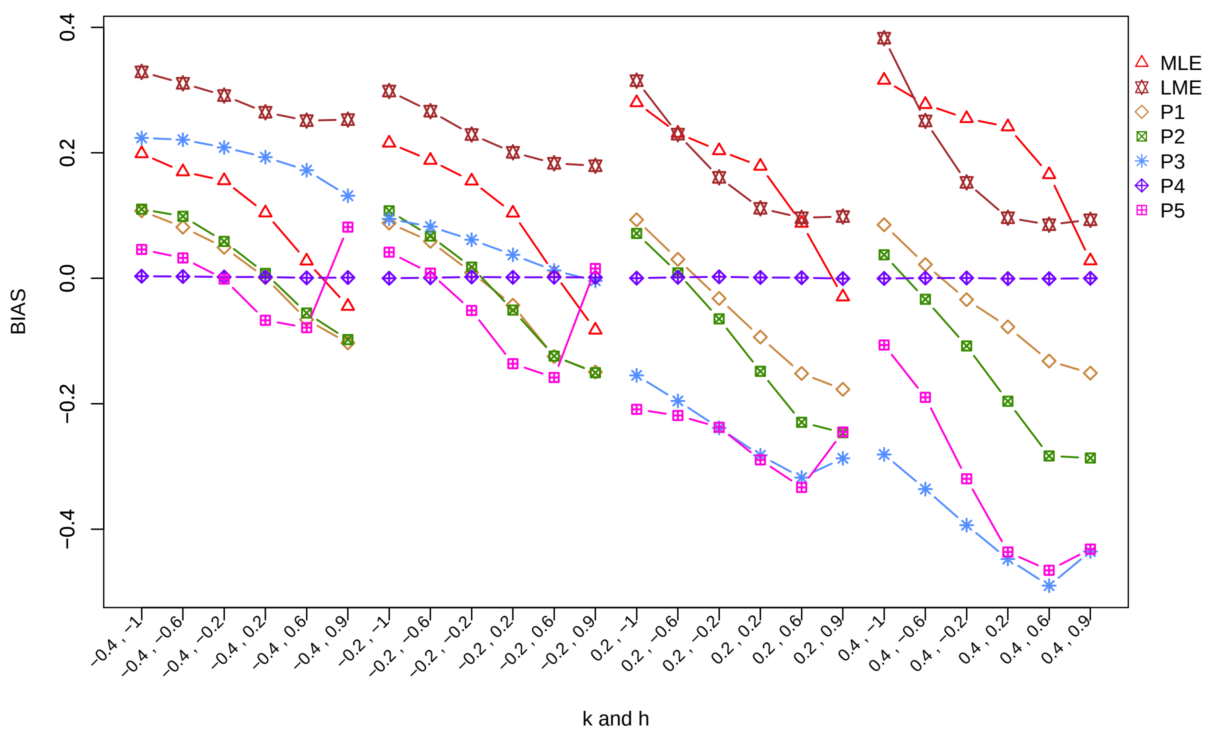

Figure S2: Average of estimates bias (BIAS) of k under various combination of k and h for sample size 30. MLE, LME and Pi stand for the method of L-moments estimator, the maximum likelihood estimator, and Bayesian estimator with prior Pi.

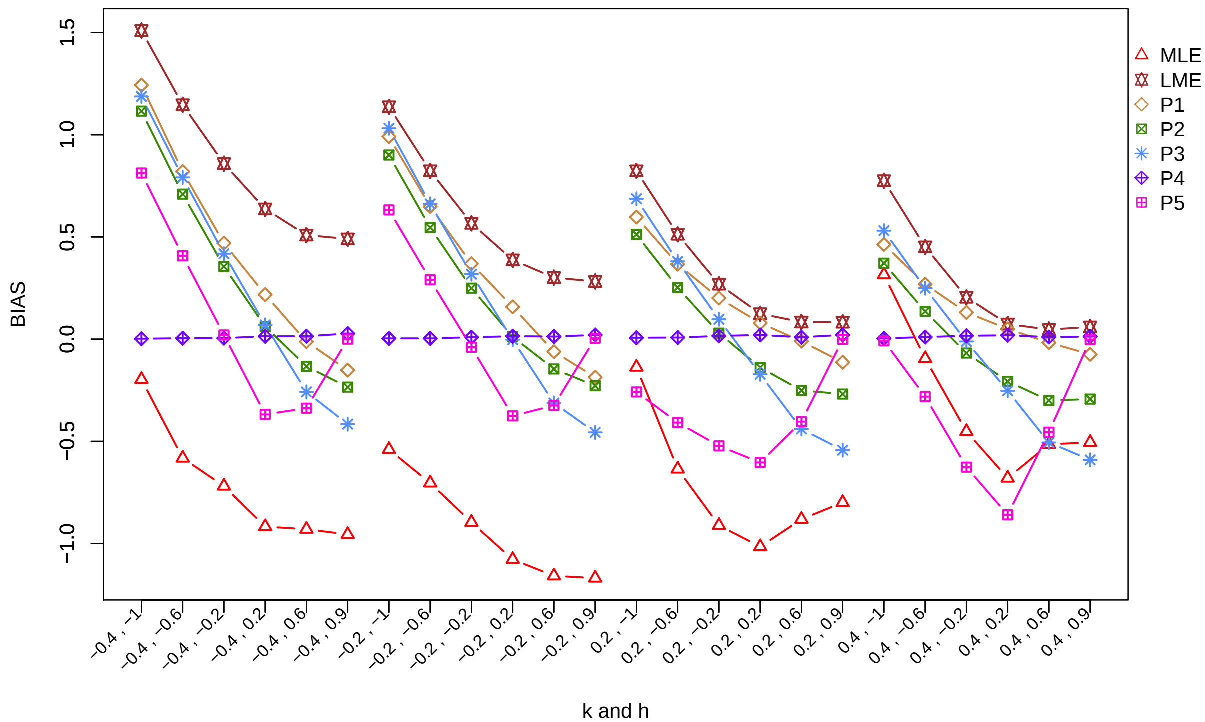

Figure S3: Same as Figure S2 but for h.

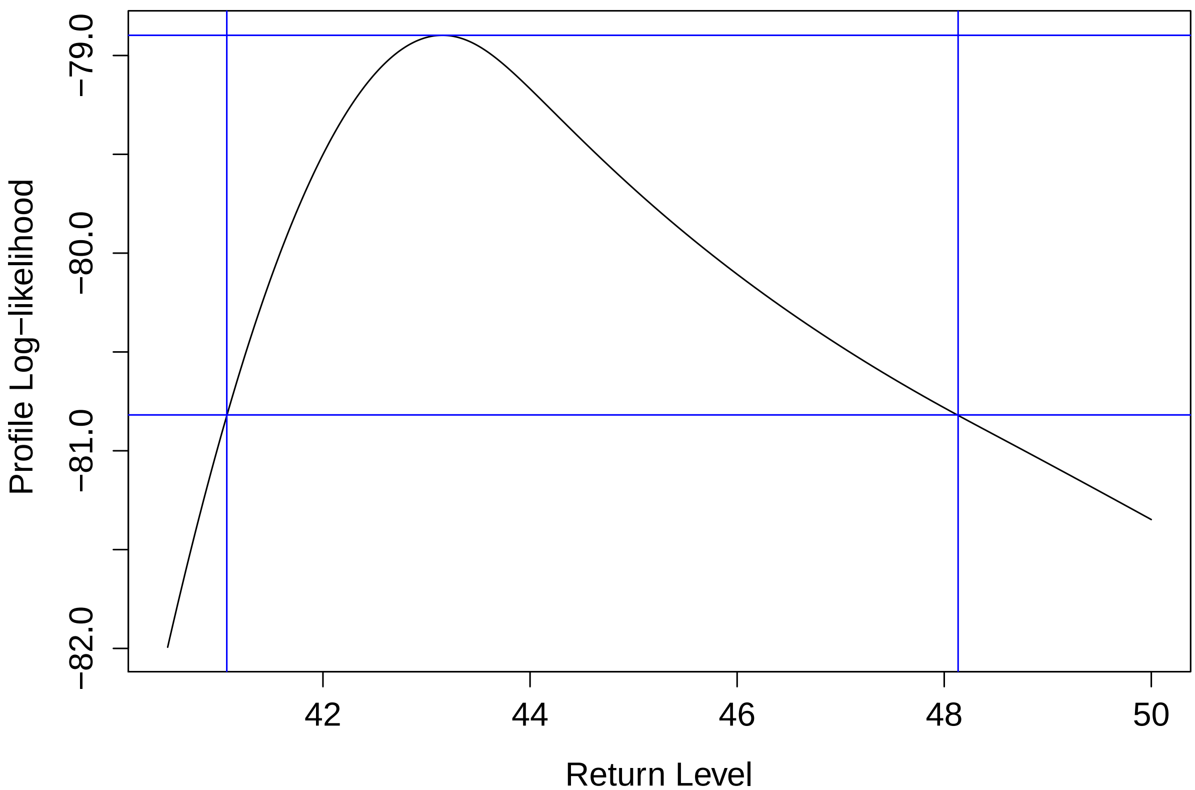

Figure S4: Profile likelihood and 95% confidence interval for 20-year return value of K4D for the annual maximum wind speeds at Udon Thani, Thailand, where parameters are estimated by maximum likelihood method.

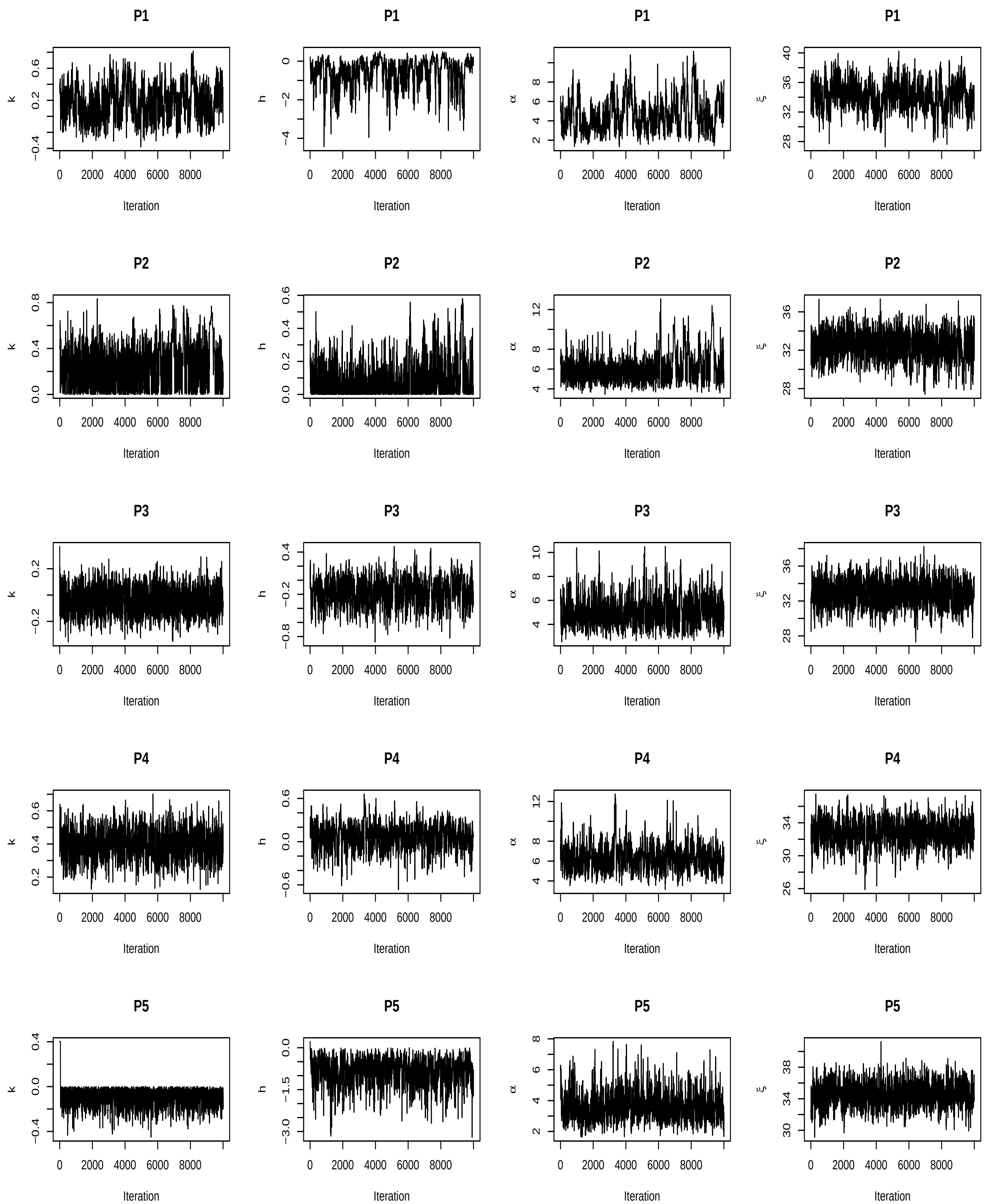

Figure S5: Trace plots of the K4D parameters using five priors for the annual maximum wind speed at Udon Thani, Thailand.

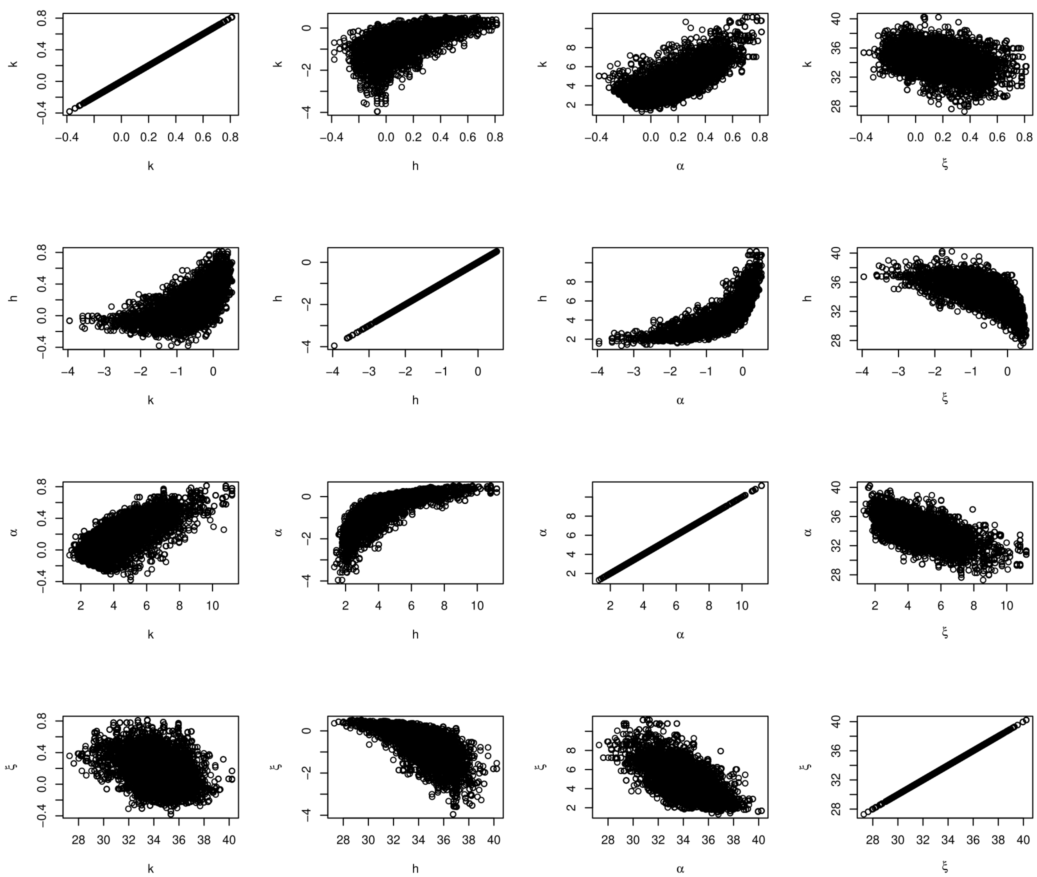

Figure S6: The posterior correlations of the parameters obtained from the annual maximum wind speed at Udon Thani, Thailand.

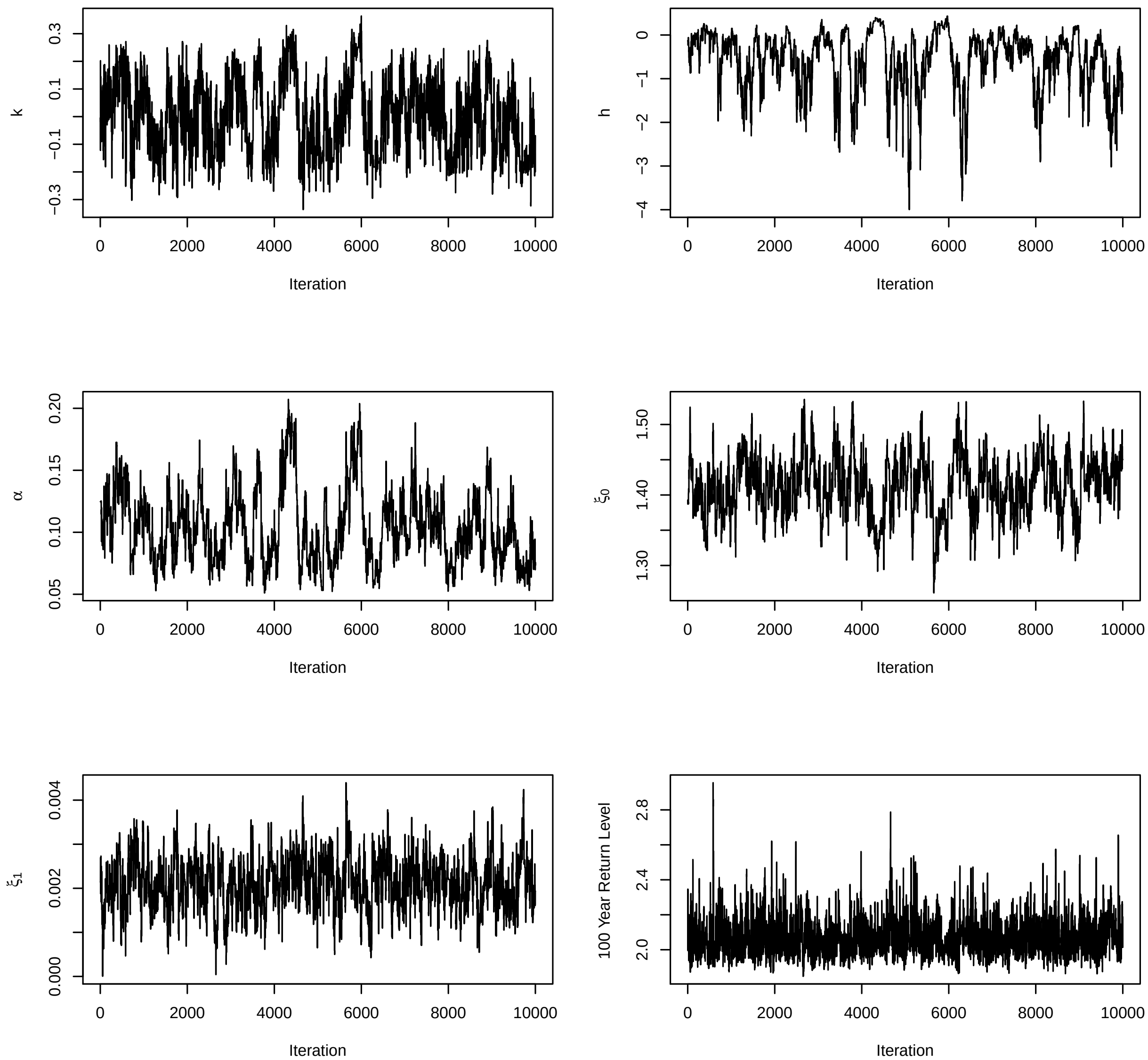

Figure S7: MCMC realizations of posterior density function and 100-year return level with the non-stationary prior P1.

References

[1] Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Texts in statistical Science. Chapman and Hall, New York, 2nd edition.

[2] Chernick, M. R. and LaBudde, R. A. (2011). *An Introduction to Bootstrap Methods with Applications to R*. John Wiley and Sons, Inc, Canada.

[3] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag London, Great Britain.

[4] Zhang, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):281–294.