

# Bayesian method for estimating Weibull parameters for wind resource assessment in the Equatorial region: a comparison between two-parameter and three-parameter Weibull distributions

M. G. M. Khan (✉ [khan\\_mg@usp.ac.fj](mailto:khan_mg@usp.ac.fj))

The University of the South Pacific

M. Rafiuddin Ahmed

The University of the South Pacific

---

## Research Article

**Keywords:** Wind energy resource assessment, Two-parameter Weibull distribution, Three-parameter Weibull distribution, Bayesian estimation method

**Posted Date:** January 11th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-504670/v2>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Bayesian method for estimating Weibull parameters for wind resource assessment in the Equatorial region: a comparison between two-parameter and three-parameter Weibull distributions

M.G.M. Khan<sup>†</sup> and M. Rafiuddin Ahmed

School of Information Technology, Engineering, Mathematics and Physics, The University of the South Pacific, Suva, Fiji.

† Corresponding author: M.G.M. Khan ([khan\\_mg@usp.ac.fj](mailto:khan_mg@usp.ac.fj))

## Abstract

The two-parameter Weibull distribution has garnered much attention in the assessment of wind energy potential. The estimation of the shape and scale parameters of the distribution has brought forth a successful tool for the wind energy industry. However, it may be inappropriate to use the two-parameter Weibull distribution to assess energy at every location, especially at sites where low wind speeds are frequent, such as the Equatorial region. In this work, a robust technique in wind resource assessment using a Bayesian approach for estimating Weibull parameters is first proposed. Secondly, the wind resource assessment techniques using a two-parameter Weibull distribution and a three-parameter Weibull distribution which is a generalized form of two-parameter Weibull distribution are compared. Simulation studies confirm that the Bayesian approach seems a more robust technique for accurate estimation of Weibull parameters. The research is conducted using data from seven sites in Equatorial region from 1° N of Equator to 19° South of Equator. Results reveal that a three-parameter Weibull distribution with non-zero shift parameter is a better fit for wind data having a higher percentage of low wind speeds (0-1 m/s) and low skewness. However, wind data with a smaller percentage of low wind speeds and high skewness showed better results with a two-parameter distribution that is a special case of three-parameter Weibull distribution with zero shift parameter. The results also demonstrate that the proposed Bayesian approach and application of a three-parameter Weibull distribution are extremely useful in accurate estimate of wind power and annual energy production.

**Keywords:** Wind energy resource assessment; Two-parameter Weibull distribution; Three-parameter Weibull distribution; Bayesian estimation method.

## Nomenclature

2-p	two-parameter Weibull distribution
3-p	three-parameter Weibull distribution
A	scale parameter of Weibull distribution, m/s
AD	Anderson-Darling test
AGL	above ground level
AIC	Akaike information criteria
BAYESIAN	Bayesian estimation method
BIC	Bayesian information criteria
COE	coefficient of efficiency
DIC	deviance information criteria
$\Gamma$	gamma function
$\gamma_1$	skewness of the wind speed

$k$	shape parameter of Weibull distribution, dimensionless
KS	Kolmogorov-Smirnov test
log-likelihood	log-likelihood method for the goodness of fit
MAE	mean absolute error
MAPE	mean absolute percentage error
MCMC	Markov Chain Monte Carlo method
MLE	maximum likelihood estimation method
RMSE	root mean square error, m/s
$n$	number of observations performed
$R^2$	correlation coefficient
$U$	wind speed, m/s
$\bar{U}$	mean wind speed, m/s
$\rho$	air density, kg/m <sup>3</sup>
$\theta$	shift or location parameter, m/s
$s$	standard deviation of wind speed, m/s

## 1 Introduction

Wind energy has now become one of the world's fastest-growing sources of energy. It is an inexhaustible source of energy with increasing utilization all around the world. Growing climate change concerns have prompted many developed and developing countries to implement policies that reduce their reliance on non-renewable energy and instead utilize renewable sources such as wind, hydro, and solar energies (Mostafaeipour et al., 2014). However, developing countries encounter several challenges in generating sustainable wind energy. There is a need for reliable wind data and proper assessments of a country's wind energy potential before initiating energy generation projects that would help them meet the sustainable development goals set by the United Nations.

While climate change is being experienced globally, some regions are getting affected more than the others. Pacific islands countries (PICs), particularly those in the warmer Equatorial region, are more susceptible to its effects. The contribution of the PICs to the current global greenhouse gas emissions, according to the UN Permanent Forum on Indigenous Issues (2015), is below 0.03%; yet they are among the first to be affected. It is projected that the people of PICs will be among the first who will need to adapt to climate change or be required to relocate from or abandon their traditional homeland. Some islands are already facing the impacts of climate change on their communities, infrastructure, water supply, coastal and forest ecosystems, fisheries, agriculture, and human health. Island states such as Kiribati, Marshall Islands, Tokelau and Tuvalu are the immediate victims of this phenomenon due to rising sea level. Knowledge of the effects of climate change on PICs should act as a driving force behind the commitment to decrease greenhouse emissions. The PICs, which currently depend heavily on imported fossil fuels and their by-products, need to become more energy efficient and self-reliant (Mohanty, 2012). PICs have some of the lowest rates of access to electricity and the prices of electricity are among the highest in the world due to their heavy reliance on high-cost diesel-based generation. Energy security and low-cost energy are becoming increasingly important within the region, which requires increasing investments in renewable energy technologies. PICs are also some of those most vulnerable to

natural disasters. The energy sector can be highly vulnerable to such events, which requires adequate attention to these issues in the design of energy production and distribution infrastructure. This can only be achieved by adopting resilient renewable energy policies. Most of the countries in the region have their national sustainable development plans to achieve United Nations' sustainable development goals (SDGs); for examples, the country Cook Islands, aims to have 100% renewable power generation in near future and Fiji is committed to reducing 30% of its national greenhouse emissions and achieve 99% renewable energy generation by 2030 (VNR Report, 2019).

However, lack of reliable and accurate wind resource data acts as a barrier to a clean energy future in the PICs, especially in the smaller developing islands (Kidmo et al., 2015). So far, wind resource assessment has received only limited attention in the PICs, and there is a need for further wind data collection and analysis and accurate wind energy potential assessment. World Bank provides support to PICs through the Sustainable Energy Industry Development Project (SEIDP). In various phases of renewable energy resource mapping, they support the countries to carry out an assessment of solar and wind potential. The objectives of this component are to enhance awareness and knowledge of the potential for renewable technologies (solar and wind) to the governments, power utilities and private sector, and to provide governments with a spatial planning framework to guide investments in the renewable energy sector (PPA 2015).

The utilization of wind power technology to generate energy is slowly increasing in PICs such as New Caledonia, Fiji, Vanuatu, Cook Islands and Samoa. However, there have been little to no attempts to establish wind power in many of these countries. The University of the South Pacific installed towers of 34 m height, named as Integrated Renewable Energy Resource Assessment Systems (IRERAS), in Kiribati, Nauru, Niue, Tuvalu, Tokelau, Samoa, Tonga, Fiji, Vanuatu, Solomon and Cook Islands to collect data on wind and solar energies (Gosai, 2014).

In recent years, the Weibull distribution has become a widely accepted tool in determining the potential of wind energy (Indhumathy, 2014). Wind energy professionals in different parts of the world have widely employed the use of Weibull distributions in the statistical analyses of wind characteristics and wind power density (Corotis et al., 1978). The Weibull shape parameter defines the width of wind distribution. A higher shape parameter indicates that the distribution is narrower, and the peak value is higher. The Weibull scale parameter controls the abscissa scale of the data distribution plot (Chang, 2011). Thus, the Weibull distribution function is comprehensively used for analyzing the wind power potential at a site.

Past researchers have found the two-parameter Weibull distribution to be a useful and practical tool for wind energy estimation. The advantages of two-parameter Weibull distribution include its flexibility, simplicity in parameter estimation, ability to use goodness-of-fit tests on these parameters as well as its dependence only on two parameters that can be expressed in closed form.

However, some authors suggested that the distribution is not suited for all wind regimes encountered in nature such as regimes with a high percentage of low wind speeds and bimodal distributions. Therefore, its usage cannot be generalized. To minimize errors, a suitable probability density function must be carefully selected for different wind regimes (Carta et al. 2009; Sukkiramathi and Seshaiyah, 2020).

Tuller and Brett (1984) proposed a three-parameter Weibull function in wind analysis and found that it showed better fitness and flexibility than the two-parameter Weibull function. Recently, some authors utilized the three-parameter Weibull distribution and found that it has more flexibility with improved fitness than the two-parameter Weibull distribution in wind energy assessments. Wais (2017) compared the two and three-parameter Weibull distribution to study the most appropriate distribution of wind speed. The results revealed that methods other than the three-parameter Weibull distribution cannot account for cases where the frequency of low wind speed is higher. The author compared the wind speeds for three different sites and found that the three-parameter Weibull distribution performed best when there was a greater frequency of lower wind speeds. Sukkiramathi and Seshaiyah (2020) also utilized the three-parameter Weibull distribution for analyzing wind power potential. However, to date, only limited research has been carried out on wind analysis using the three-parameter Weibull distribution.

Furthermore, many estimation methods have been proposed for estimating Weibull parameters. Among these, maximum likelihood estimation (MLE), a popular frequentist technique, has been a widely used method for estimating the parameters (Teimouri and Gupta, 2013). Recently, the Bayesian estimation approach has received great attention from many researchers. Among them is Al Omari and Ibrahim (2011), who considered the Bayesian survival estimator for Weibull distribution with censored data. Many authors, including Hossain and Zimmer (2003) and Pandey et al. (2011), did some comparative studies on the estimation of the Weibull parameters using complete and censored samples, and Lye et al. (1993) determined the Bayes estimation for the extreme-value reliability function. Guure et al. (2012) examined the performance of MLE and Bayesian methods for estimating the two-parameter Weibull failure time distribution. However, the use of the Bayesian technique for modelling wind data and analyzing wind power potential was not explored in their work.

The present work aims to compare the two-parameter and three-parameter Weibull distributions to fit wind speed data more accurately at seven locations in the Equatorial region, where wind speeds are generally lower. The aim is also to develop a novel approach using the Bayesian method for estimating the Weibull parameters. The results from Bayesian technique will be compared with those of the traditional MLE method to determine a more accurate evaluation method of wind speed characteristics.

## 2 Wind speed data

Wind speed data from seven different sites in the Equatorial region were used in the present work, as shown in Table 1.

Table 1: Locations of data collection sites

Sites	Location	Country	Measurement period	Topography
1. Tarawa	Latitude 1° 26' N Longitude 173° 00' E	Kiribati	September 2012 to September 2013	Flat
2. Pentecost	Latitude 15° 41' S Longitude 168° 11' E	Vanuatu	October 2012 to November 2013	Mountainous terrain
3. Rakiraki	Latitude 17° 22' S Longitude 178° 10' E	Fiji	February 2012 to October 2013	Flat

4. Kadavu	Latitude 19° 0' S Longitude 178° 15' E	Fiji	January 2018 to December 2018	Mountainous terrain
5. Rarotonga	Latitude 21° 15' S, Longitude 159° 45' W	Cook Islands	January 2016 to December 2018	Flat
6. Nuku'alofa	Latitude 21° 15' S Longitude 175° 15' W	Tonga	January 2016 to August 2019	Flat
7. Sanasana	Latitude 18° 6' S Longitude 177° 20' E	Fiji	December 2018 to December 2019	Flat

For sites 1, 2 and 3, data were obtained from measurements using 34 m tall towers with the help of sensors described in Table 2. The NRG systems towers, named Integrated Renewable Energy Resource Assessment Systems (IRERAS), with a height of 34 m were used. NRG *SymphoniePlus3* was the data-logger used which was connected to seven different sensors installed on the tower. The sensors measured wind speed, temperature, pressure, rainfall, solar insolation, humidity, and wind direction. The data were either collected from the SD card in person or sent via the GSM-based network to a data-bank located at the ICT centre of the University of South Pacific at the Laucala Campus, Fiji. The anemometers (serial numbers 179500189054-57, 179500189089-90) have an accuracy of 0.1 m/s and a range of 0.4 to 96 m/s. The wind vane is placed at 30 m AGL. The data were recorded in a time-series format in an RWD file which were later transferred to a Microsoft excel sheet. The wind speed data were recorded continuously at an interval of 10 minutes with a cup anemometer at hub heights of 34 m and 20 m, respectively. For sites 4, 5 and 6, satellite data were downloaded; land data from ERA5 were used in the present work (Ref: <https://cds.climate.copernicus.eu/>). ERA5 is the fifth generation ECMWF reanalysis for the global climate and weather for the past 4 to 7 decades. Reanalysis combines model data with observations from across the world into a globally complete and consistent dataset using the laws of physics. This principle, called data assimilation, is based on the method used by numerical weather prediction centres, after certain number of hours (12 hours at ECMWF) a previous forecast is combined with newly available observations in an optimal way to produce a new best estimate of the state of the atmosphere, called analysis, from which an updated, improved forecast is issued. Reanalysis works in the same way, but at reduced resolution to allow for the provision of a dataset spanning back several decades. Reanalysis does not have the constraint of issuing timely forecasts, so there is more time to collect observations, and when going further back in time, to allow for the ingestion of improved versions of the original observations, which all benefit the quality of the reanalysis product. For site 7, an NRG systems towers, similar to the ones used for sites 1, 2 and 3, but 50 m high, was installed; it has anemometers at 50 m, 40 m and 30 m AGL.

Table 2: Specifications of the measurement sensors (Aukitino et al., 2017)

Parameter	Sensor Type	Range	Accuracy
<b>Wind speed</b>	NRG#40C anemometer	0.4-96.0 m/s	0.1 m/s
<b>Wind direction</b>	NRG 200P direction vane	0-360°	N/A
<b>Pressure</b>	NRG BP-20 barometric pressure sensor	15 kPa - 115 kPa	1.5 kPa
<b>Temperature</b>	NRG 110S	-40 °C – 65 °C	1.11 °C

For the measured values, some uncertainties were taken into account such as calibration errors, the terrain of the site that was used, the dynamic over speeding, the error introduced due to wind shear and the inflow angle (Jain, 2016). The measurements in the present work were performed close to the shoreline at a flat terrain. The flow was in the horizontal plane, resulting in a lower uncertainty level. The calibration report for the anemometers used in the present work showed a maximum uncertainty of 0.6% for a wind speed range of 4-7 m/s, which reduced at higher wind speeds. The overall uncertainty in the estimation of wind speed is obtained by taking all the above uncertainties into account (Jain, 2016) and using the relation in equation (1):

$$\varepsilon = \sqrt{\sum_{i=1}^N \varepsilon_i^2} \quad (1)$$

where  $\varepsilon_i$  is each component of uncertainty and  $N$  is the number of components of uncertainty. The uncertainties were estimated at 95% confidence level. As per the IEC Standard IEC 61400-12-1 (IEC, 2017), the uncertainty in the measurements was estimated to be approximately 1.74%.

### 3 Weibull Distribution

Assessment of wind power energy at a site requires knowledge of the appropriate probability distribution of the site's wind speed, as the estimation of wind energy depends on its accuracy.

#### 3.1 Two-parameter Weibull distribution

The two-parameter Weibull probability density functions (pdf) and the cumulative distribution function (cdf) for wind speed ( $U$ ), respectively are given by

$$f(U) = \frac{k}{A} \left( \frac{U}{A} \right)^{k-1} e^{-\left(\frac{U}{A}\right)^k}; \quad U > 0, k > 0, A > 0 \quad (2)$$

and

$$F(U) = 1 - e^{-\left(\frac{U}{A}\right)^k} \quad (3)$$

where  $f(U)$  is the probability of observing the wind speed,  $k$  is the shape parameter and  $A$  is the scale parameter (m/s) of the distribution. The parameter  $k$  indicates the wind potential and what peak the distribution can reach. Its value ranges between 1 and 3. A lower  $k$  value signifies highly variable winds, while constant winds are characterized by a larger  $k$ . The parameter  $A$  denotes how windy the site under study is and it takes a value proportional to the mean wind speed (Manwell et al., 2010; Sukkiramathi and Seshaiyah, 2020).

#### 3.2 Three-parameter Weibull distribution

The three-parameter Weibull pdf and the cdf for wind speed, respectively are given by

$$f(U) = \frac{k}{A} \left( \frac{U - \theta}{A} \right)^{k-1} e^{-\left(\frac{U - \theta}{A}\right)^k}; \quad U > 0, k > 0, A > 1, -\infty < \theta < \infty \quad (4)$$

and

$$F(U) = 1 - e^{-\left(\frac{U-\theta}{A}\right)^k} \quad (5)$$

where  $f(U)$  is the probability of observing the wind speed,  $k$  is the shape parameter,  $A$  is the scale parameter (m/s), and  $\theta$  is the shift or location parameter (m/s) of the distribution. If  $\theta = 0$ ,  $f(U)$  and  $F(U)$  become the PDF and CDF of a two-parameter Weibull distribution, respectively.

As the name implies, the shift parameter,  $\theta$ , shifts the distribution along the abscissa. When  $\theta = 0$ , the distribution starts at  $U = 0$  or at the origin. Whereas, if  $\theta > 0$ , the distribution starts at the location  $\theta$  to the right of the origin. If  $\theta < 0$ , the distribution starts at the location  $\theta$  to the left of the origin. For the distribution of wind speed,  $\theta$  provides an estimate of the earliest time-to-start the wind (Tuller and Brett, 1984; Wais, 2017).

## 4 Methods of Estimating Weibull Parameters

To estimate the Weibull parameters, we propose a Bayesian approach and compare its performance with a popular frequentist approach, the maximum likelihood estimation (MLE) method.

### 4.1 The Maximum Likelihood Method

#### 4.1.1 Two-parameter distribution

MLE is the most popular technique for deriving estimators (Casella and Berger, 2002; Aukitino et al., 2017; Chaurasiya et al., 2018). If  $U_1, \dots, U_n$  are the wind speed values with the Weibull density function given in (2), the shape parameter ( $k$ ) and scale parameter ( $A$ ) are the values that maximize the likelihood function  $L(k, A | U_1, \dots, U_n) = \prod_{i=1}^n f(U_i | k, A)$ . Then, solving  $\partial \ln L / \partial k = 0$  and  $\partial \ln L / \partial A = 0$  gives the equation of MLE of the scale parameter  $A$  as:

$$A = \frac{1}{n} \sum_{i=1}^n U_i^k \quad (6)$$

Finally, equation (7) is used for estimating the shape parameter ( $k$ ) as:

$$\frac{1}{k} + \frac{1}{n} \sum_{i=1}^n \ln(U_i) - \frac{\sum_{i=1}^n U_i^k \ln(U_i)}{\sum_{i=1}^n U_i^k} = 0, \quad (7)$$

which may be solved to obtain the estimate of  $k$  using Newton-Raphson method or any other numerical procedure because equation (7) does not have a closed form solution. When  $k$  is obtained, the value of  $A$  is found from equation (6).

### 4.1.2 Three-parameter distribution

The likelihood function  $L$  for estimating parameters is given by  $L(k, A, \theta | U_1, \dots, U_n) = \prod_{i=1}^n f(U_i | k, A, \theta)$ . Then, solving  $\partial \ln L / \partial k = 0$ ,  $\partial \ln L / \partial A = 0$  and  $\partial \ln L / \partial \theta = 0$  gives the equations of MLE of the parameters as shown in Equations (8-10):

$$\frac{n}{k} \sum_{i=1}^n \log\left(\frac{U_i - \theta}{A}\right) - \sum_{i=1}^n \left(\frac{U_i - \theta}{A}\right)^k \log\left(\frac{U_i - \theta}{A}\right) = 0 \quad (8)$$

$$\frac{nk}{A} + \frac{k}{A} \sum_{i=1}^n \left(\frac{U_i - \theta}{A}\right)^k = 0 \quad (9)$$

$$-(k-1) \sum_{i=1}^n \frac{1}{U_i - \theta} - \frac{k}{A} \sum_{i=1}^n \left(\frac{U_i - \theta}{A}\right)^{k-1} = 0 \quad (10)$$

There is no closed form solution of the equations but the non-linear equations (8) - (10) may be solved by applying some optimization techniques such as Newton-Raphson method or other numerical procedures (Lawless, 2003; Teimouri and Gupta, 2013).

### 4.1.3 Evaluation of MLE methods

To determine the best model, we can compare the fit of the two MLE methods using different measures of goodness-of-fit (Luceño, 2008; Ramachandran and Tsokos, 2015; Cousineau and Allan, 2015). The most used criteria are:

#### Log-likelihood (log-like):

If a pdf  $f_{\hat{\eta}}(U)$  fitted on the wind speed data and  $\hat{\eta}$  is the estimated parameter of the distribution, then the log-likelihood for the goodness of fit is obtained by the following equation:

$$\text{log-like} = \log\left(\prod_{i=1}^n f_{\hat{\eta}}(U_i)\right) \quad (11)$$

where  $U_i$  is the  $i$ th observed wind speed and  $n$  is the number of observations in the dataset. A higher value of log-likelihood value indicates a better fit.

#### Akaike information criteria (AIC):

If  $k$  is the number of distribution parameters to estimate, the AIC is obtained using equation (12):

$$\text{AIC} = -2(\text{log-like}) + 2k \quad (12)$$

A lower value of AIC indicates that the model fits the data better. Compared to the log-likelihood, this criterion takes into consideration the parsimony of the model as it includes a penalty term that increases the number of parameters.

**Bayesian information criteria (BIC):**

This criterion is obtained using equation (13):

$$\text{BIC} = -2(\log\text{-like}) + k \log(n) \quad (13)$$

Similar to AIC, a lower value of BIC indicates that the model fits the data better. However, BIC provides a stronger penalty than AIC for additional parameters.

**Kolmogorov-Smirnov (KS):**

The Kolmogorov-Smirnov (KS) test is also used to check the adequacy of a given theoretical distribution for a given set of wind speed data. The KS test computes the maximum difference between the predicted and observed distribution, and the test statistic  $D$  is given by:

$$D = \max_{1 \leq i \leq n} |F_i - \hat{F}_i| \quad (14)$$

where  $\hat{F}_i$  is the  $i$ th predicted cumulative probability from the theoretical cdf and  $F_i$  is the empirical probability of the  $i$ th observed wind speed.

**Anderson-Darling (AD):**

For a finite data sample, the Anderson-Darling (AD) test statistic  $A^2$  is defined by:

$$A^2 = -n - s, \quad (15)$$

where  $s = \sum_{i=1}^n \frac{2i-1}{n} \left[ \log(\hat{F}_i) + \log(1 - \hat{F}_{n-i+1}) \right]$ .

**4.2 The Bayesian Method**

A classical frequentist approach such as MLE has certain drawbacks. Most of its properties hold only for large sample sizes and it requires a symmetric form of sample distribution. The Bayesian approach, however, is free from such limitations. Moreover, Bayesian simulation tools provide an exact method of inference even if sample size is very small. Thus, in real-life situations where the sample size may be small, Bayesian methods seem to be more suitable over frequentist methods if prior information about the parameters is available.

In this paper, a Bayesian inference approach for modeling of wind speed data is proposed. In the Bayesian paradigm, data and prior information about the parameters are combined together to make an inference about the parameters of interest.

The most influential contribution of the Bayesian approach is its modification of the likelihood function into a posterior - a valid probability distribution defined by the classic Bayes' rule. The posterior distribution of wind speed is expressed as:

$$p(\eta|U) = \frac{p(U|\eta)p(\eta)}{p(U)} \quad (16)$$

where  $p(\eta|U)$  is the posterior distribution of wind speed,  $p(\eta)$  is the prior distribution of unknown parameters  $\eta = (k, A, \theta)$ ,  $p(U|\eta)$  is the likelihood of wind speed data and  $p(U) = \int p(U|\eta)p(\eta)d\eta$ . The denominator of equation (16),  $p(U)$ , normalizes the posterior distribution,  $p(\eta|U)$ . Since it is independent of  $U$ , it is often convenient to write the posterior distribution as:

$$p(\eta|U) \propto p(U|\eta)p(\eta), \quad (17)$$

that is, the posterior distribution of the parameters is proportional to the likelihood function times the prior distribution of parameters. While fitting wind speed data, we use a non-informative uniform prior distribution as we have very little prior knowledge about its model parameters. In Bayesian computations, a sample of the joint posterior distribution is obtained by using Gibbs sampler to simulate a sample from a Markov Chain Monte Carlo (MCMC). Then, we can calculate the desired values of the posterior.

In this paper, the software JAGS is used to fit the model. The R package `R2jags` is used to summarize the posterior inference, which is discussed in more detail in Section 4.2.1.

#### 4.2.1 Bayesian Fitting of Weibull distribution with JAGS

JAGS, an acronym for ‘‘Just Another Gibbs Sampler’’ (Plummer, 2003), accepts a model string written in an R-like syntax that compiles and generates MCMC samples from the model using Gibbs sampling. It is an open-source software written in C++ using GNU compilers and packaging tools - freely available at <http://mcmc-jags.sourceforge.net/>. R packages such as `R2jags` or `rjags` allow running JAGS models within R on Windows machines for the summarization of posterior inference.

In the present work, we develop JAGS models for the Bayesian fit of wind speed with two and three parameter Weibull distributions and generate MCMC data. In MCMC simulations, we run the Gibbs sampler with the JAGS models for 10,000 iterations using the `jags` function in `R2jags` package (Su and Yajima, 2020). Then, the posterior estimates of parameters are obtained by performing the Gibbs sampler iterations and using 1000 burn-in period to attain convergence with five thinning intervals and 3 chains with a sample size of 1800 per chain. The model specifications to perform the Bayesian fit, and the `jags` function and its arguments are presented in Appendix 1.

#### 4.2.2 Evaluation of Bayesian Models

The standard likelihood, AIC and BIC statistics, discussed in Section 4.1.3 are not relevant while evaluating Bayesian methods such as MCMC. Instead, Spiegelhalter et al. (2002) suggested that

the Deviance Information Criterion (DIC) be used to compare models. The DIC is a generalization of AIC that is based on Deviance statistics:

$$D(\eta) = -2\log f(U/\eta) + 2\log h(U), \quad (18)$$

where  $h(U)$  is some standardizing function of the data. The DIC is then defined as:

$$DIC = \bar{D} + p_D, \quad (19)$$

where  $\bar{D} = E_{\eta|U}(D)$  is the posterior expectation of deviance and  $p_D$  is effective number of parameters that captures the complexity of a model. A smaller value of DIC indicates a better-fitting model.

## 5 Estimate of wind power and energy

When the wind speed ( $U$ ) of a site and the frequency distribution  $f(U)$  are known, the available wind power and wind energy can be estimated. If  $\rho$  is the density of air,  $D$  is the turbine rotor diameter and  $A_R = \pi D^2/4$  is the rotor cross sectional area, then the probability of available wind power for a given velocity  $U$  is obtained by

$$p(U) = \frac{1}{2} \rho A_R U^3; \quad U > 0. \quad (20)$$

Then, the expected wind power ( $P$ ) is estimated by

$$\text{Wind power, } P = \int_0^{\infty} P(U) \cdot f(U) dU. \quad (21)$$

Substituting (20), (2) and (4) in (21), the expected wind powers for two-parameter and three-parameter Weibull distributions respectively, are determined by

$$P_{2P} = \frac{1}{2} \rho A_R \frac{k}{A^k} \int_0^{\infty} U^{k+2} e^{-\left(\frac{U}{A}\right)^k} dU \quad (22)$$

and

$$P_{3P} = \frac{1}{2} \rho A_R \frac{k}{A^k} \int_0^{\infty} U^3 \cdot (U - \theta)^{k-1} e^{-\left(\frac{U-\theta}{A}\right)^k} dU \quad (23)$$

If  $t_{\text{year}} = 8760$  is the total number of hours in a year, the expected wind energy for two-parameter and three-parameter Weibull distributions respectively, are determined by

$$E_{2P} = P_{2P} \cdot t_{\text{year}} \quad (24)$$

and

$$E_{3P} = P_{3P} \cdot t_{\text{year}} \quad (25)$$

## 6 Analyzing performance of different estimators

The efficiency and performance of MLE and Bayesian methods for estimating two and three parameter Weibull distributions were determined using different goodness of fit and error measures such as coefficient of determination ( $R^2$ ), root mean square error (RMSE), coefficient of efficiency (COE), mean absolute error (MAE) and the mean absolute percentage error (MAPE). Arithmetically, these are computed as follows (Azad et al., 2014; Kidmo et al., 2015; Aukitino et al., 2017):

### Coefficient of determination ( $R^2$ ):

It is a statistical measure that gives some information about the goodness of fit of a model, that is, how much the variance of the observed data is explained by the fitted model. It is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^n (U_i - \hat{U}_i)^2}{\sum_{i=1}^n (U_i - \bar{U})^2}, \quad (26)$$

where  $n$  is the number of observations,  $U_i$  is the  $i$ th actual data,  $\hat{U}_i$  is the  $i$ th predicted data with the Weibull distribution,  $\bar{U}$  is the mean of actual data. A higher  $R^2$  value indicates a better fit and  $R^2 = 1$  indicates that the regression predictions perfectly fit the data.

### Root mean square error (RMSE):

It determines the deviation of the predicted values of wind speed from the observed values and obtained by

$$RMSE = \left[ \frac{1}{n} \sum_{i=1}^n (U_i - \hat{U}_i)^2 \right]^{\frac{1}{2}} \quad (27)$$

A smaller RMSE value normally indicates accurate modeling. The calculated RMSE value approaches zero as the difference between the observed and predicted values becomes smaller (Indhumathy et al., 2014).

### Coefficient of efficiency (COE):

It quantifies the ratio of difference between predicted wind speed and the mean wind speed to the difference between actual values and the average of wind speeds. A higher COE value indicates a good fit for the data. It is expressed as:

$$COE = \frac{\sum_{i=1}^n (\hat{U}_i - \bar{U})^2}{\sum_{i=1}^n (U_i - \bar{U})^2}. \quad (28)$$

**Mean absolute error (MAE):**

The mean absolute error is a measure of the absolute difference between predicted and actual values. A smaller value of MAE indicates higher accuracy. The MAE is mathematically expressed as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{U}_i - U_i|. \quad (29)$$

**Mean absolute percentage error (MAPE):**

It is a comparative measure, indicating the error as a percentage of the actual data which helps accurately predict the forecasting method. Like MAE, a lower value of MAPE indicates better accuracy. It is mathematically expressed as:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{\hat{U}_i - U_i}{U_i} \right|. \quad (30)$$

**7 Results**

In this section, the results of fitting of two-parameter (2-p) and three-parameter (3-p) Weibull distributions are presented. Further, the results for the application of MLE and the proposed Bayesian approach for estimating the parameters, as discussed in Sections 3 and 4, are also presented. To accomplish this, wind speed data at seven different sites were collected as mentioned in Section 2. Table 3 provides wind speed distributions at these sites. The table shows that the range of speed varies at different sites. The lowest range of wind speed was observed at site 1 (0-19 m/s) and the highest range was found at site 3 (0-34 m/s). Some sites tend to have more low to null speed (0-1 m/s).

Table 3: Frequency distributions of wind speed at different sites

Wind speed $U$	Site 1 Frequency	Site 2 Frequency	Site 3 Frequency	Site 4 Frequency	Site 5 Frequency	Site 6 Frequency	Site 7 Frequency
(0,1]	889	1063	5882	62	209	251	1662
(1,2]	2197	1998	4246	219	773	799	2161
(2,3]	5224	3579	4951	367	1498	1543	3712
(3,4]	7006	5739	5845	641	1928	2390	5220
(4,5]	8721	8077	7134	803	2388	3330	6331
(5,6]	8783	10332	8716	949	2878	4260	7393
(6,7]	7554	9570	10771	1026	2965	4590	6698
(7,8]	5591	7094	11215	1052	2937	3915	5651
(8,9]	3381	4810	9363	851	2833	3502	4449
(9,10]	1706	2701	6496	882	2349	2812	3535
(10,11]	604	1255	4226	723	2016	2023	2792
(11,12]	236	482	2366	437	1411	1205	1989
(12,13]	125	119	1177	311	892	715	1316
(13,14]	42	26	588	215	661	278	654
(14,15]	11	9	247	99	405	193	288
(15,16]	3	8	107	54	91	114	67
(16,17]	1	10	60	7	42	97	9

(17,18]	4	3	53	9	17	43	1
(18,19]	2	3	38	6	6	30	0
(19,20]	0	1	13	6	2	16	0
(20,21]	0	0	5	7	3	9	0
(21,22]	0	0	4	15	0	8	0
(22,23]	0	0	3	8	0	11	0
(23,24]	0	0	6	2	0	1	0
(24,25]	0	0	5	5	0	1	0
(25,26]	0	0	4	2	0	0	0
(26,27]	0	0	4	0	0	0	0
(27,28]	0	0	3	1	0	0	0
(28,29]	0	0	5	0	0	0	0
(29,30]	0	0	10	0	0	0	0
Above 30	0	0	4	1	0	0	0
Total frequency	52080	56879	83554	8760	26304	32136	53928

Both the 2-p and 3-p Weibull distributions were fitted to the recorded wind speed data and the parameters in the distributions were estimated using the MLE and the Bayesian methods. In the MLE method, the goodness of fit with 2-p and 3-p Weibull distributions is evaluated using the statistical measures AIC, BIC, AD, KS and log-like. Table 4 presents the estimated values of the parameters of 2-p and 3-p Weibull distributions and the values of the various statistical measures determined by the MLE at the seven sites.

Table 4: MLE estimated values of parameters and statistical measures

Site	Distribution	$k$	$A$	$\theta$	AIC	BIC	AD	KS	log-like
1	2-p Weibull	2.564486	6.019568		230352.8	230370.6	23.95999	0.021077	-115174.4
	3-p Weibull	2.777792	6.438856	-0.380611	<b>230075.7</b>	<b>230102.3</b>	<b>9.98482</b>	<b>0.013953</b>	<b>-115034.9</b>
2	2-p Weibull	2.735978	6.535565		256112.4	256130.3	101.4772	0.038265	-128054.2
	3-p Weibull	3.233794	7.532297	-0.922046	<b>255252.9</b>	<b>255279.8</b>	<b>32.0876</b>	<b>0.025796</b>	<b>-127623.5</b>
3	2-p Weibull	1.948074	7.017269		433669.2	433687.8	1414.721	0.092759	-216832.6
	3-p Weibull	2.636074	8.804806	-1.546749	<b>429666.9</b>	<b>429694.9</b>	<b>478.748</b>	<b>0.058134</b>	<b>-214830.5</b>
4	2-p Weibull	2.385540	8.445309		45434.68	45448.83	6.759304	0.019735	-22715.34
	3-p Weibull	2.401323	8.493057	-0.042536	45435.82	45457.05	6.577964	0.019739	<b>-22714.91</b>
5	2-p Weibull	2.438715	8.234657		135020.5	135036.9	16.29118	0.015911	-67508.26
	3-p Weibull	2.569510	8.596043	-0.323893	<b>134971.8</b>	<b>134996.4</b>	<b>10.37910</b>	<b>0.014681</b>	<b>-67482.92</b>
6	2-p Weibull	2.503177	7.843472		159616.8	159633.6	22.04708	0.019279	-79806.41
	3-p Weibull	2.522584	7.896873	-0.047855	<b>159615.8</b>	159641.0	<b>21.85474</b>	<b>0.018862</b>	<b>-79804.91</b>
7	2-p Weibull	2.153354	7.112349		271584.1	271601.9	69.32922	0.021085	-135790.1
	3-p Weibull	2.425388	7.807676	-0.603869	<b>270987.3</b>	<b>271014.0</b>	<b>20.57445</b>	<b>0.010117</b>	<b>-135490.6</b>

In Bayesian estimates, uniform prior distributions of the parameters were used to fit wind data. Firstly, a sample of the joint posterior distribution by simulating a sample from MCMC methods using a Gibbs sampler as discussed in Section 4.2.1 is obtained. Finally, the DIC is obtained to evaluate the Bayesian parameters of the Weibull distributions. Table 5 presents the estimated mean values of the parameters with standard deviation (SD) of both two and three parameter Weibull distributions for all the sites. The 95% credible region (lower limit - 2.5% and upper limit - 97.5%) for each of the parameters and the model evaluation statistic DIC values are also presented.

Table 5: Estimated values of the parameters using obtained using Bayesian technique and summary statistics

Site	Distribution	Parameter	Mean	SD	2.5%	97.5%	DIC
1	2-p Weibull	k	2.564487	0.008762	2.547292	2.581560	949864.6
		A	6.019417	0.010876	5.998086	6.041268	
	3-p Weibull	k	2.778195	0.018610	2.742305	2.814923	<b>949587.6</b>
		A	6.439670	0.033898	6.373026	6.507357	
		$\theta$	-0.381144	0.029685	<b>-0.439400</b>	<b>-0.323193</b>	
2	2 parameter	k	2.736076	0.008940	2.718410	2.753285	256112.2
		A	6.535458	0.010427	6.515002	6.555774	
	3 parameter	k	3.231943	0.023096	3.184961	3.273617	<b>255252.4</b>
		A	7.528328	0.044253	7.438378	7.607035	
		$\theta$	-0.918508	0.040458	<b>-0.989749</b>	<b>-0.835880</b>	
	3	2 parameter	k	1.947640	0.005556	1.936812	1.958478
A			7.017671	0.013022	6.992180	7.042895	
3 parameter		k	2.635850	0.016118	2.604492	2.667974	<b>429666.8</b>
		A	8.804653	0.044092	8.718409	8.892835	
		$\theta$	-1.546999	0.038766	<b>-1.624995</b>	<b>-1.472661</b>	
4	2-p Weibull	k	2.385596	0.019486	2.347390	2.424420	<b>166458.7</b>
		A	8.446062	0.039889	8.367275	8.525195	
	3-p Weibull	k	2.405809	0.026990	2.356128	2.461114	166459.8
		A	8.507469	0.070222	8.377540	8.660753	
		$\theta$	-0.055714	0.051466	<b>-0.169674</b>	<b>0.031573</b>	
	5	2-p Weibull	k	2.438528	0.011898	2.415595	2.462772
A			8.233447	0.021645	8.190758	8.276122	
3-p Weibull		k	2.569184	0.024930	2.521912	2.619417	<b>498375.1</b>
		A	8.596434	0.065046	8.471752	8.729620	
		$\theta$	-0.324597	0.055076	<b>-0.438669</b>	<b>-0.220035</b>	
6	2-p Weibull	k	2.502775	0.010741	2.482126	2.524369	603592.3
		A	7.843795	0.018777	7.806656	7.880416	
	3-p Weibull	k	2.523740	0.016243	2.492982	2.555752	<b>603591.3</b>
		A	7.900740	0.038140	7.830501	7.978555	
		$\theta$	-0.051032	0.030016	<b>-0.113036</b>	<b>0.004279</b>	
7	2-p Weibull	k	2.153266	0.007378	2.139086	2.167660	1016627
		A	7.111621	0.014636	7.083240	7.140584	
	3-p Weibull	k	2.425984	0.016281	2.394208	2.457741	<b>1016030</b>
		A	7.808854	0.040671	7.730066	7.892452	
		$\theta$	-0.604597	0.033817	<b>-0.673024</b>	<b>-0.539979</b>	

The goodness-of-fit criteria and summary statistics presented in Tables 4 and 5 indicate that the **3-p Weibull** distribution fits better than the 2-p Weibull distribution for wind speed at the sites 1, 2, 3, 5 and 7 as all the goodness-of-fit measures (AIC, BIC, AD, KS and log-likelihood) are smaller in MLE estimate and the DIC is also smaller in Bayesian estimate. Moreover, as shown in Table 5 and the marginal posterior means of shift parameter  $\theta$  are -0.381144, -0.918508, -1.546999, -0.324597 and -0.604597. The 95% credible regions of this parameter as shown in columns 2.5% and 97.5% clearly indicates that the **3-p Weibull** distribution is more appropriate at these sites as the value of its shift parameter is non-zero (i.e.  $\theta \neq 0$ ).

However, for the site 4, both MLE results in Tables 4 and 5 indicate that the **2-p Weibull** distribution fits better than the 3-p Weibull distribution at this site as the AIC, BIC and KS are smaller in MLE estimate and the DIC is smaller in Bayesian estimate. Although, the marginal

posterior mean of  $\theta$  of 3-p Weibull distribution is -0.055714 but its 95% credible region (-0.169674, 0.031573) indicates that the value of its shift parameter is zero (i.e.  $\theta = 0$ ). This implies that 3-p Weibull reduces to the 2-p Weibull indicating that **2-p Weibull** distribution is more appropriate at this site.

Whereas, at the site 6, the results show a lack of significant difference between the 2-p and 3-p Weibull distributions while fitting wind speeds, as all MLE and Bayesian estimates are similar in numerical value. Also, the 95% credible region indicates that  $\theta = 0$  in Bayesian estimate, which indicates **2-p Weibull** may be a better distribution at this site as the 3-p Weibull distribution reduces to the 2-p Weibull distribution. It should be noted that the 3-p Weibull distribution is a generalized form of 2-p distribution, and when  $\theta = 0$ , it becomes 2-p distribution. Thus, 2-p distribution can be seen as a special case of 3-p distribution.

Finally, since the Bayesian estimation is a simulation based iterative procedure, the convergence of the model is diagnosed by the visual inspections using the trace and posterior density plots. A sample of such plots obtained for the parameters of 2-p Weibull and 3-p Weibull distributions from the Bayesian simulations are presented in Figures 1-4 for sites 3 and 4. The plots for the remaining sites are presented in Figures 12-21 in Appendix 2.

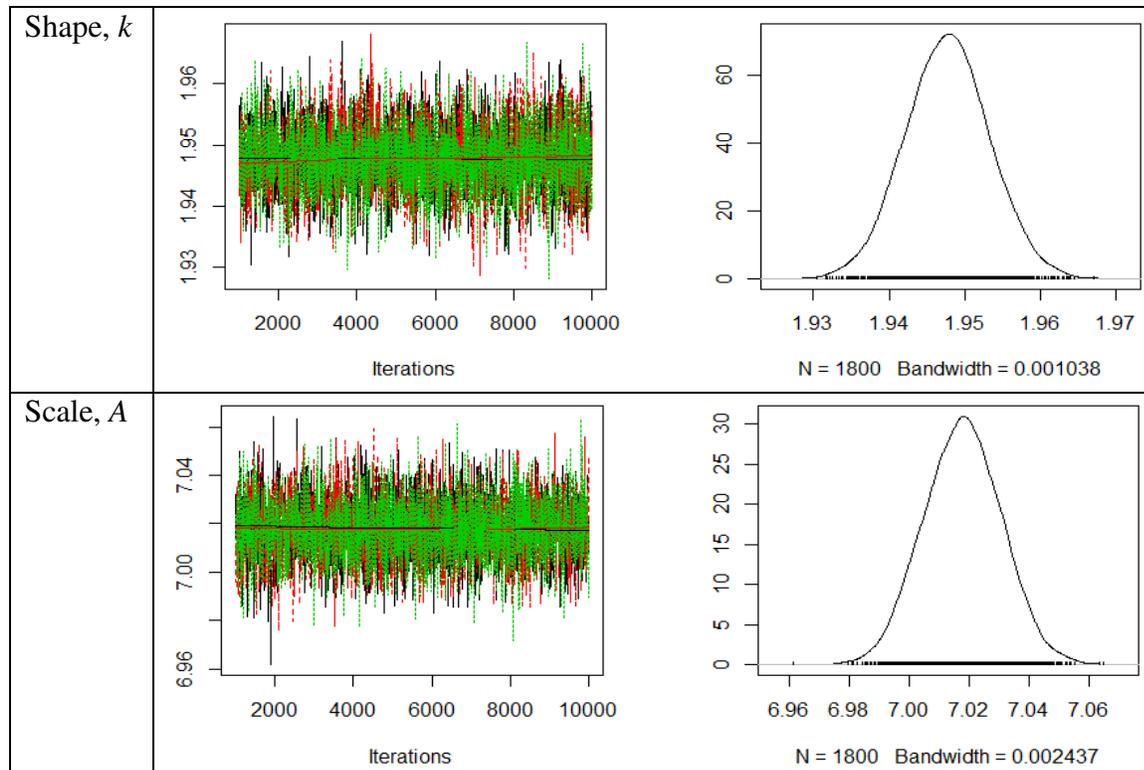


Figure 1: Trace and posterior density plots for site 3 (2-p Weibull). Trace plot (left) shows a ‘fat hairy caterpillar’ appearance which is indicating a random scatter around the stable mean of the shape parameter  $k = 1.947640$  within 95% credible region (1.936812, 1.958478) and the scale parameter  $A = 7.017671$  within 95% credible region (6.992180, 7.042895). The density plots (right) also display smooth curves of these simulated values for both the parameters. Thus, the

plots clearly indicate the convergence of the simulations of the Bayesian estimates presented in Table 5.

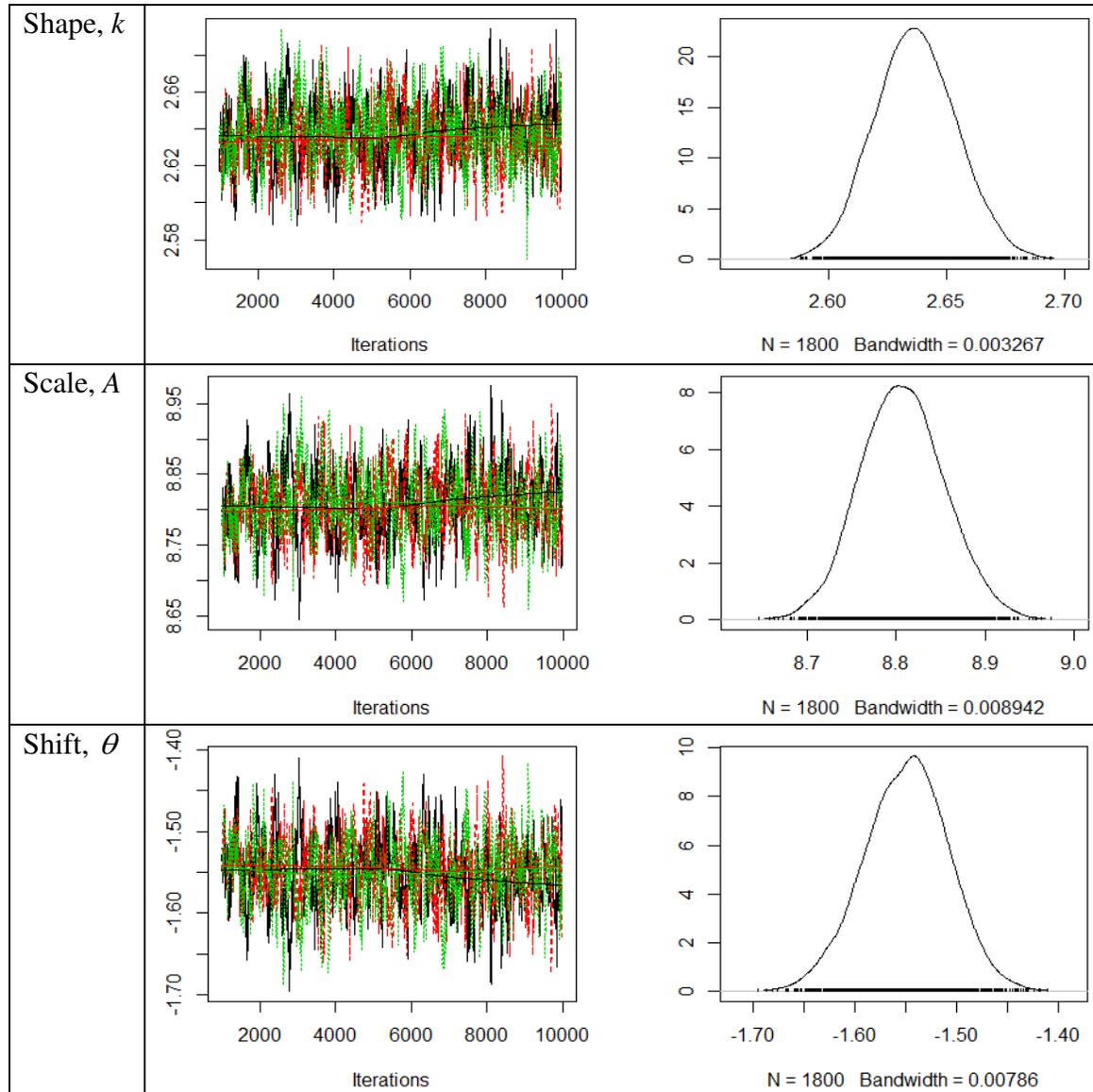


Figure 2: Trace and posterior density plots for site 3 (3-p Weibull). Trace plot (left) shows a ‘fat hairy caterpillar’ appearance which is indicating a random scatter around the stable mean of the shape parameter  $k = 2.635850$  within 95% credible region (2.604492, 2.667974) and the scale parameter  $A = 8.804653$  within 95% credible region (8.718409, 8.892835). The shift parameter also shows a random scatter around the stable mean of  $\theta = -1.546999$  within 95% credible region (-1.624995, -1.472661), which is far away from zero and it reveals the appropriateness of using 3-p Weibull. The density plots (right) also display smooth curves of these simulated values for all the parameters. Thus, the plots clearly indicate the convergence of the simulations of Bayesian estimates presented in Table 5.

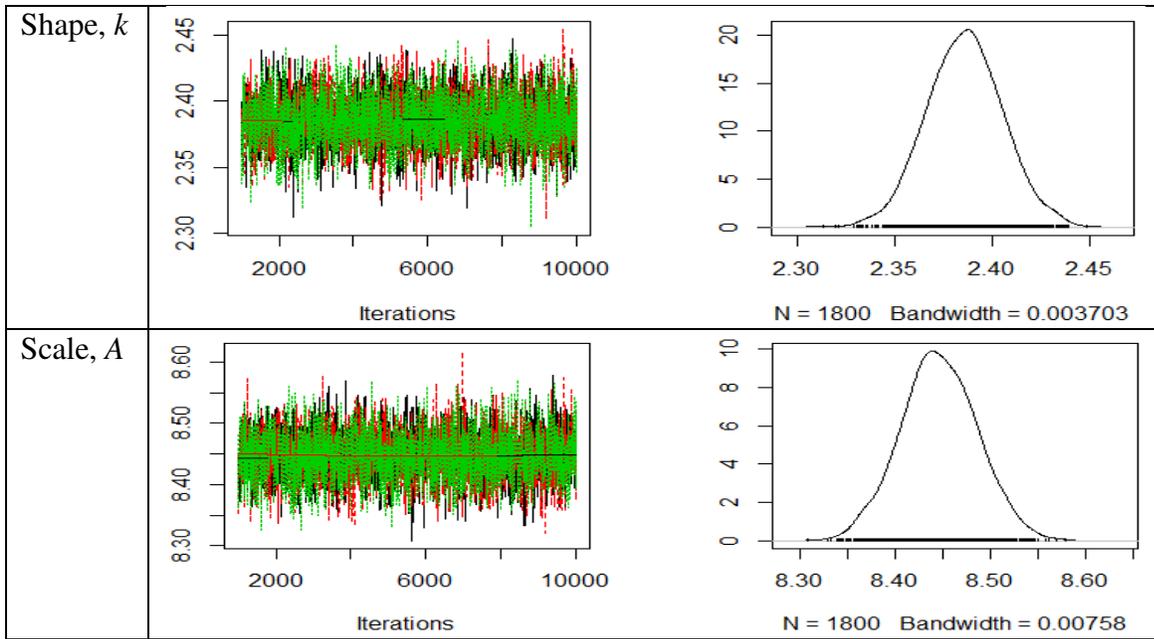
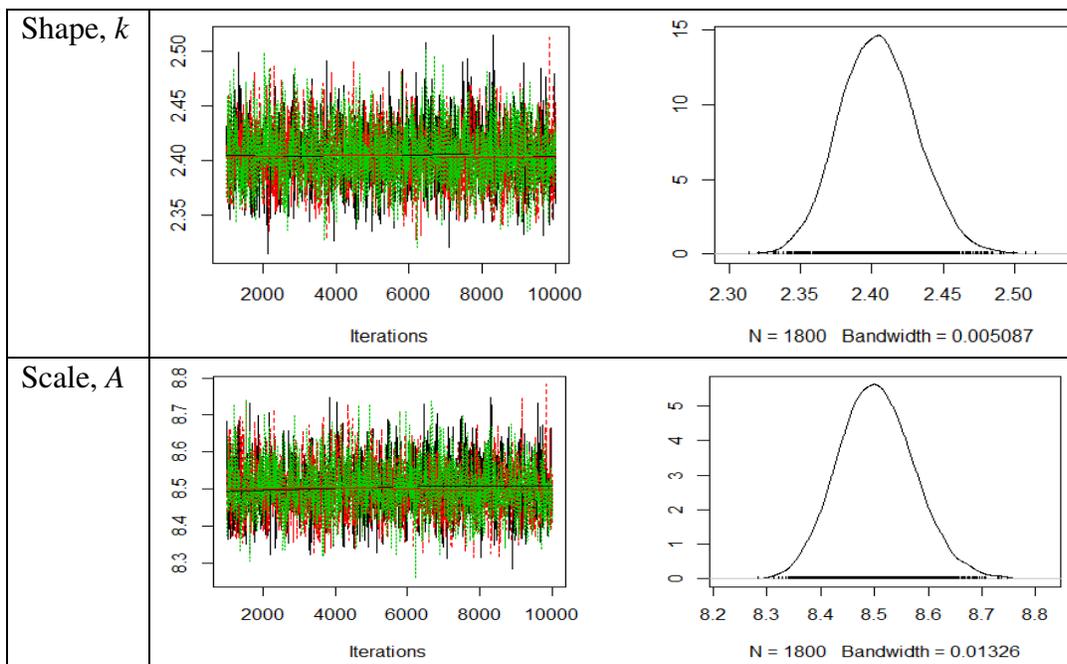


Figure 3: Trace and posterior density plots for site 4 (2-p Weibull). Trace plot (left) shows a ‘fat hairy caterpillar’ appearance which is indicating a random scatter around the stable mean of the shape parameter  $k = 2.385596$  within 95% credible region (2.347390, 2.424420) and the scale parameter  $A = 8.446062$  within 95% credible region (8.367275, 8.525195). The density plots (right) also display smooth curves of these simulated values for both the parameters. Thus, the plots clearly indicate the convergence of the simulations of Bayesian estimates presented in Table 5.



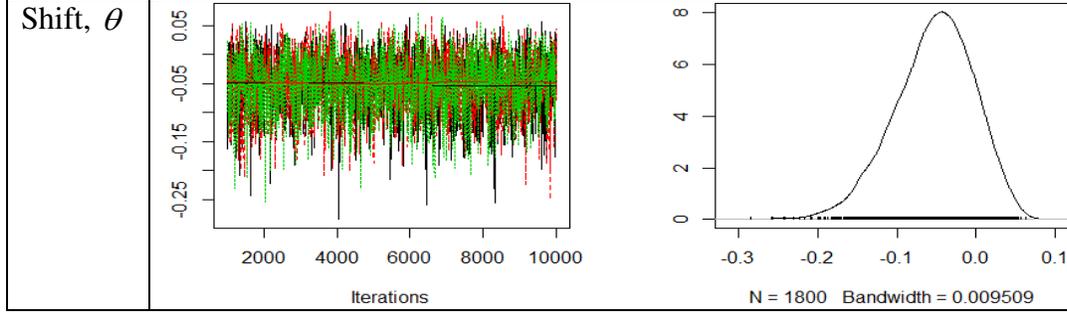


Figure 4: Trace and posterior density plots for site 4 (3-p Weibull). Trace plot (left) shows a ‘fat hairy caterpillar’ appearance which is indicating a random scatter around the stable mean of the shape parameter  $k = 2.405809$  within 95% credible region (2.356128, 2.461114) and the scale parameter  $A = 8.507469$  within 95% credible region (8.377540, 8.660753). The shift parameter also shows a random scatter around the stable mean of  $\theta = -0.055714$  within 95% credible region (-0.169674, 0.031573), which includes zero and it reveals the appropriateness of using 2-p Weibull. The density plots (right) also display smooth curves of these simulated values for all the parameters. Thus, the plots clearly indicate the convergence of the simulations of Bayesian estimates presented in Table 5.

## 8 Discussion

In Section 7, the results for the goodness of fit for the wind speed distributions at seven different sites were presented. Results showed that the 3-p Weibull distribution was a better fit for wind speeds at all the sites investigated, except the sites 4 and 6, in the Equatorial region. However, the 2-p Weibull distribution may be a better fit for wind speeds data at the sites 4 and 6 as the shift parameter  $\theta$  in 3-p Weibull was found to be zero as detected in Bayesian estimate. As discussed earlier, the 2-p Weibull distribution can be considered a special case of 3-p Weibull distribution.

In this section, further investigations were carried out to explain the difference between the performance of the two distributions. Referring to the percentage of lowest wind speed (0-1 m/s) presented in Table 6, the results clearly show that the wind distributions of the sites (sites 1, 2, 3, 5 and 7) that have high percentage (0.79% - 7.04%) of lower (or closer to null) wind speeds perfectly fit the 3-p Weibull distribution. The shift parameter in the simulations is also found to be significant i.e.  $\theta \neq 0$  for these sites. On the other hand, the 2-p Weibull distribution was a better fit for wind speed distributions at Sites 4 and 6, where the percentage of low wind speed was smaller ( $< 0.78\%$ ). Similar findings were reported by Wais (2017).

Moreover, histograms presented in Figures 5-11 for wind distributions at sites 1 to 7 show different shapes, indicating a variation in skewness. Thus, another reason for fitting a better distribution is the skewness of the wind speed distribution. The skewness ( $\gamma_1$ ) is a measure of the asymmetry of the wind speed distribution about its mean, which is defined for a sample of  $n$  values as:

$$\gamma_1 = \frac{m_3}{s^3} \quad (31)$$

where,  $s$  = standard deviation and  $m_3 = \frac{1}{n} \sum_{i=1}^n (U - \bar{U})^3$  is the third central moment. A normal distribution is symmetrical and has  $\gamma_1 = 0$ . If  $\gamma_1$  is negative, the distribution is left skewed whereas a positive  $\gamma_1$  indicates a right skewed distribution. Since the Weibull distribution is a right skewed one,  $\gamma_1$  is expected to be positive.

Table 6 presents the mean ( $\bar{U}$ ), standard deviation ( $s$ ) and skewness ( $\gamma_1$ ) of the wind speed data at each site. It shows that the wind speed distributions of sites 4 and 6 have higher skewness compared to sites 1, 2, 3, 5 and 7.

**Table 6: Percentage lowest wind speed, mean, SD and best Weibull distribution**

Site	% lowest speed (0-1 m/s)	Site mean wind speed ( $\bar{U}$ )	SD ( $s$ )	Skewness ( $\gamma_1$ )	Fitted distribution
1	1.71	5.35	2.22	0.28	3-p Weibull
2	1.87	5.83	2.29	0.09	3-p Weibull
3	7.04	6.29	3.20	0.20	3-p Weibull
4	0.71	7.50	3.31	0.62	2-p Weibull
5	0.79	7.31	3.20	0.22	3-p Weibull
6	0.78	6.97	2.95	0.51	2-p Weibull
7	3.08	6.32	3.04	0.29	3-p Weibull

Thus, the results reveal that the 3-p Weibull distribution is a better fit for wind speed data with both: greater frequency of low wind speeds (0-1 m/s) and low skewness, compared to a 2-p Weibull distribution. The Bayesian analysis also confirms that the wind speed data with a smaller percentage of low speeds fitted better as a 2-p Weibull distribution than the 3-p Weibull distribution as its location parameter  $\theta$  was found to be zero.

To reiterate, this research is aimed at comparing the goodness of fit of both 2-p and 3-p Weibull distributions and to compare the performance of frequentist MLE and Bayesian methods for the estimation of Weibull parameters. Therefore, a comparison study of the four methods is conducted by:

1. Fitting of 2-p Weibull distribution with an MLE estimate – MLE.2p
2. Fitting of 3-p Weibull distribution with an MLE estimate – MLE.3p
3. Fitting of 2-p Weibull distribution with a Bayesian estimate – BAYESIAN.2p
4. Fitting of 3-p Weibull distribution with a Bayesian estimate – BAYESIAN.3p

The wind speed bins and the Weibull distribution curves obtained using these four methods are illustrated in Figures 5-11. The figures show that for all the sites, except the sites 4 and 6, the 3-p Weibull distribution are closer to the histograms than the 2-p Weibull distributions. The Weibull distribution curves of Bayesian estimates are also closer to the histograms than MLE estimates. Thus, the fitting of 3-p Weibull distribution for the wind speed data is better than the fitting using the 2-p Weibull distribution method (Aukitino et al., 2017).

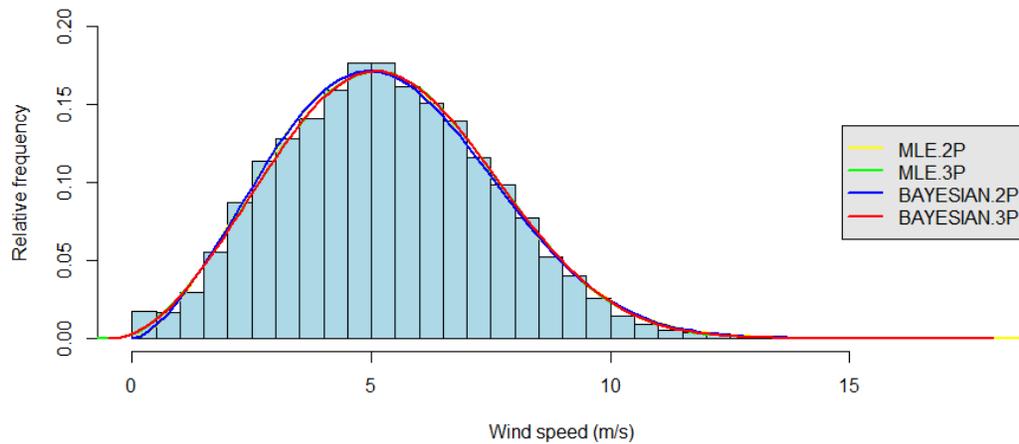


Figure 5: 2-p and 3-p Weibull curves by four methods and histogram of the observed wind speeds at site 1. It can be seen that the wind speed distribution is almost symmetric and the standard deviation of the site's wind speed is the lowest compared to the other sites. The mean wind speed is close to the centre; the site's mean wind speed is 5.35 m/s. The percentage of lowest wind speed (0 – 1 m/s) is relatively high at this site, hence the Bayesian 3-p Weibull distribution fitted the data best as can be seen from Tables 5 and Table 7.

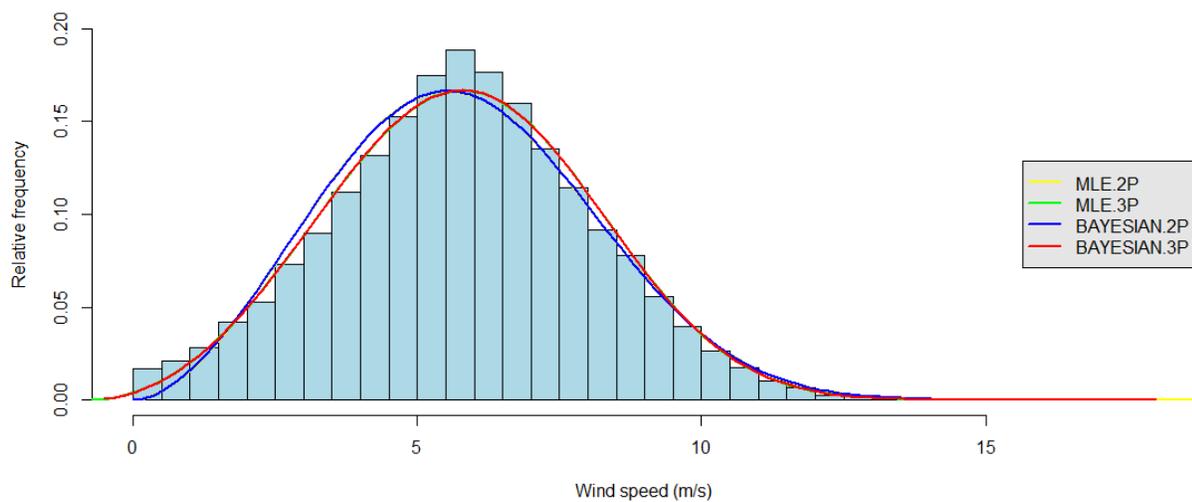


Figure 6: 2-p and 3-p Weibull curves by four methods and histogram of the observed wind speeds at site 2. Very symmetric wind speed bars on both the sides of the peak which is 6-7 m/s and hence a symmetric Weibull distribution curve for this site with a very small skewness can be seen. The percentage of lowest wind speeds of 0-1 m/s is similar to site 1, while the mean wind speed is 5.83 m/s, which is evident from a comparison of Figures 5 and 6. It can also be seen that the Bayesian 3-p distribution fits the histogram well. It is clear from Table 7 that  $R^2$  is the highest, while the RMSE, MAE and MAPE values are the lowest for the Bayesian 3-p method at this site.

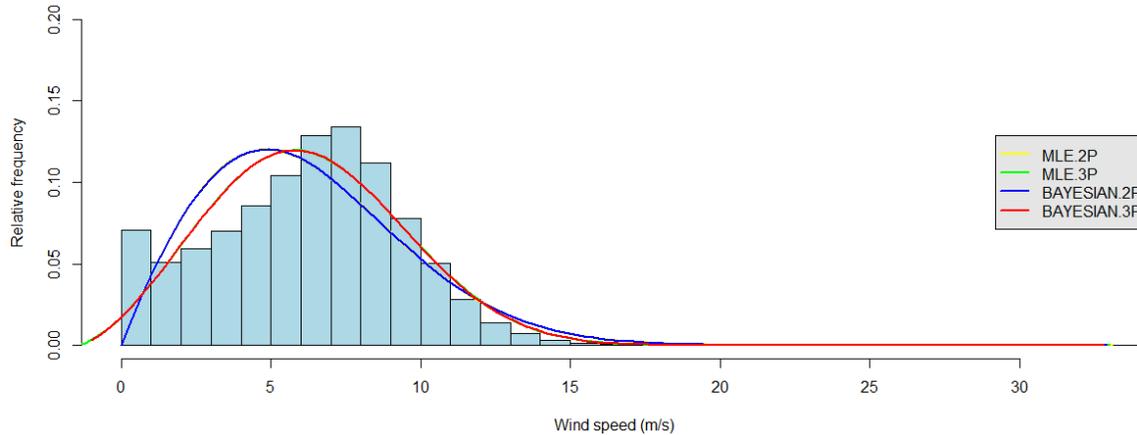


Figure 7: 2-p and 3-p Weibull curves by four methods and histogram of the observed wind speeds at site 3. The figure shows a non-symmetric distribution of wind speeds with the highest percentage of wind speeds of 0-1 m/s and a higher standard deviation. The mean wind speed for this site is 6.29 m/s. Due to the high occurrence of very low wind speeds, the Bayesian 3-p Weibull distribution is a better fit for this case. However, due to the peculiar distribution of wind speeds at this site, the Weibull distribution does not fit the histogram perfectly like the previous two sites. It is clear from Table 7 that  $R^2$  is the highest, while the RMSE, MAE and MAPE values are the lowest for the Bayesian 3-p method for this site.

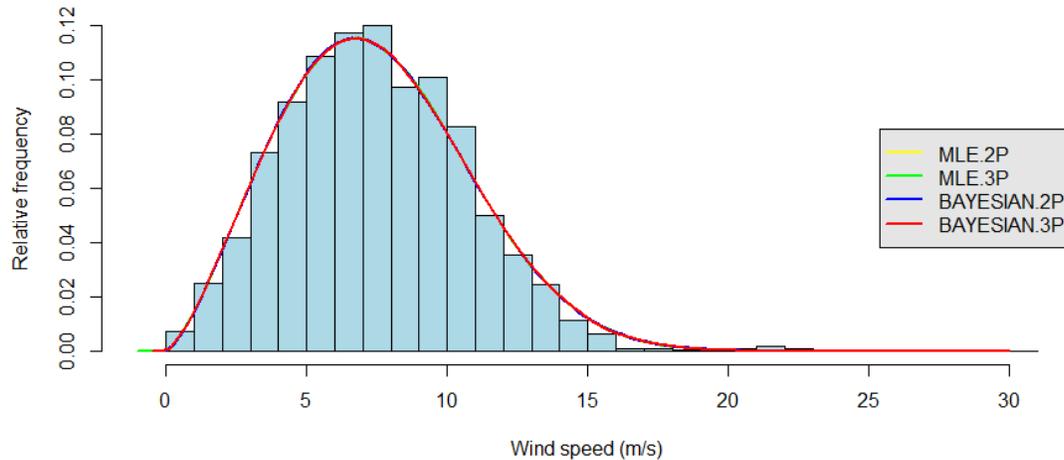


Figure 8: 2-p and 3-p Weibull curves by four methods and histogram of the observed wind speeds at site 4. The figure shows a relatively less symmetric wind speed distribution for site 4 compared to sites 1 and 2 with a clearly high skewness. For this case, the percentage of the lowest wind speeds in the range of 0-1 m/s is the lowest, hence the Bayesian 2-p distribution fits the wind speed data well. Although the four Weibull distribution curves seem to be coinciding, the Bayesian 2-p distribution curve lies above the 3-p one before the peak and below the 3-p curve after the peak. From Table 4, it can be seen that the value of the shift parameter  $\theta$  is the smallest (close to zero) for this site, hence the 3-p Weibull distribution becomes 2-p distribution.

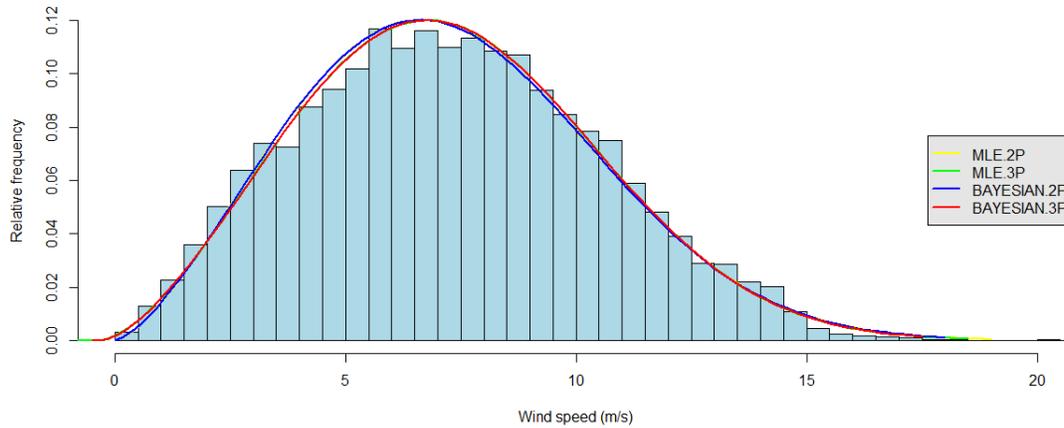


Figure 9: 2-p and 3-p Weibull curves by four methods and histogram of the observed wind speeds at site 5. There are relatively higher frequencies of lower and higher wind speeds compared to the mean wind speed. However, the skewness is still relatively low for this site. The percentage of lowest wind speed of 0-1 m/s is 0.79% and it can be seen that the Bayesian 3-p Weibull distribution fits the histogram better. The fit is supported by the analysis presented in Tables 5 and 7.

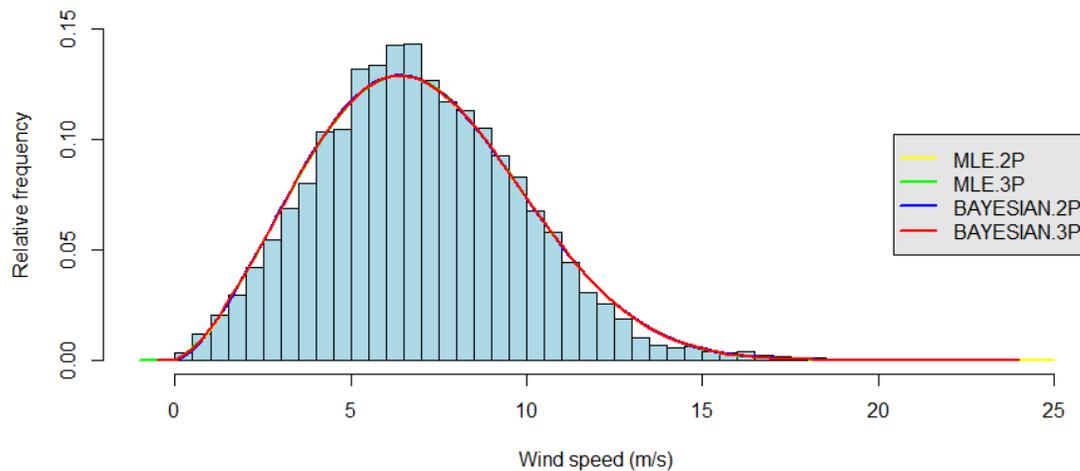


Figure 10: 2-p and 3-p Weibull curves by four methods and histogram of the observed wind speeds at site 6. The skewness is relatively higher for this site. The mean wind speed is close to 7 m/s and the standard deviation is relatively higher. The percentage of the smallest wind speeds of 0-1 m/s is 0.78% for which the 2-p Weibull distribution is a better fit as per the analysis in Tables 5 and 7.

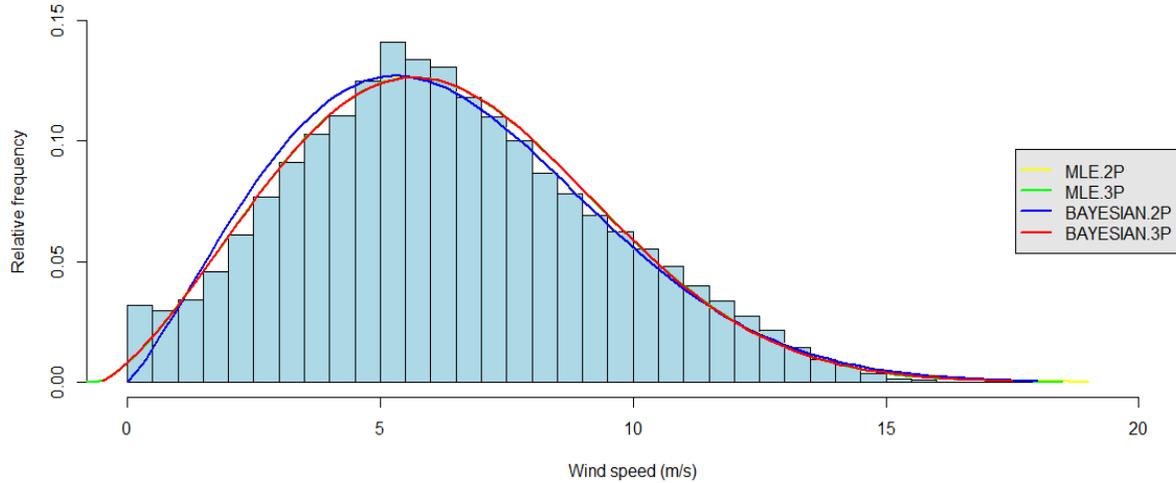


Figure 11: 2-p and 3-p Weibull curves by four methods and histogram of the observed wind speeds at site 7. This site has the second highest percentage of smallest wind speeds. It can be seen that the Bayesian 3-p distribution is a better fit here with the 3-p curve touching almost all the bars at the peak. The error analysis of Table 7 clearly shows that the 3-p Bayesian curve is the best fit for this site’s wind speed data.

To evaluate the performance of the four methods, the five statistical goodness of fit measures discussed in Section 6:  $R^2$ , RMSE, COE, MAE and MAPE are estimated and compared. The results of the estimation of these measures for the wind data from all the seven sites are presented in Table 7, which clearly reveal that BAYESIAN.3P is the most efficient method for the sites 2, 3, 5 and 7 as it produces highest  $R^2$  value and the least RMSE, MAE and MAPE values. This indicates that the 3-p Weibull with Bayesian estimates is a better method for wind energy assessment at these sites. Moreover, the method equally performs with MLE.3p for the site 1. Whereas, BAYESIAN.2P performs better for the sites 4 and 6 producing highest  $R^2$  value and the least COE and RMSE values, which indicates 2-p Weibull with Bayesian estimates is a better method for assessment at these two sites which have the lowest occurrence of smallest wind speeds in the range of 0-1 m/s.

Table 7: Goodness-of-fit measures for four methods

Site	Method	$R^2$	COE	RMSE	MAE	MAPE
1	MLE.2P	0.9982	<b>0.9936</b>	0.0983	0.0421	1.3052
	MLE.3P	<b>0.9983</b>	1.0011	<b>0.0975</b>	<b>0.0239</b>	<b>0.8269</b>
	BAYESIAN.2P	0.9982	<b>0.9936</b>	0.0983	0.0421	1.3053
	BAYESIAN.3P	<b>0.9983</b>	1.0011	<b>0.0975</b>	<b>0.0239</b>	<b>0.8269</b>
2	MLE.2P	0.9949	1.0048	0.1655	0.1114	3.4259
	MLE.3P	<b>0.9971</b>	<b>1.0034</b>	0.1234	0.0661	1.7917
	BAYESIAN.2P	0.9949	1.0048	0.1655	0.1114	3.4259
	BAYESIAN.3P	<b>0.9971</b>	1.0041	<b>0.1227</b>	<b>0.0645</b>	<b>1.7748</b>
3	MLE.2P	0.9696	0.9823	0.5675	0.4596	10.7507
	MLE.3P	<b>0.9755</b>	1.0681	0.5219	0.3029	7.9253
	BAYESIAN.2P	0.9696	<b>0.9822</b>	0.5677	0.4597	10.7516
	BAYESIAN.3P	<b>0.9755</b>	1.0679	<b>0.5216</b>	<b>0.3027</b>	<b>7.9149</b>

4	MLE.2P	<b>0.9787</b>	1.0512	0.4983	0.1546	1.9778
	MLE.3P	0.9783	1.0523	0.5025	0.1522	1.9013
	BAYESIAN.2P	<b>0.9787</b>	<b>1.0511</b>	<b>0.4981</b>	0.1544	1.9724
	BAYESIAN.3P	0.9783	1.0524	0.5036	<b>0.1518</b>	<b>1.8860</b>
5	MLE.2P	0.9983	<b>1.0180</b>	0.1371	0.1031	2.0079
	MLE.3P	0.9987	1.0186	0.1109	0.0787	1.5343
	BAYESIAN.2P	0.9983	1.0181	0.1372	0.1033	2.0074
	BAYESIAN.3P	<b>0.9989</b>	1.0183	<b>0.1108</b>	<b>0.0785</b>	<b>1.5295</b>
6	MLE.2P	<b>0.9911</b>	1.0173	0.2809	<b>0.1251</b>	1.9635
	MLE.3P	0.9909	1.0181	0.2852	0.1257	1.9187
	BAYESIAN.2P	<b>0.9911</b>	<b>1.0169</b>	<b>0.2806</b>	0.1253	1.9673
	BAYESIAN.3P	0.9909	1.0180	0.2854	0.1258	<b>1.9162</b>
7	MLE.2P	0.9973	<b>0.9991</b>	0.1607	0.1126	3.4174
	MLE.3P	0.9987	1.0152	0.1069	0.0718	2.0482
	BAYESIAN.2P	0.9973	0.9992	0.1609	0.1130	3.4244
	BAYESIAN.3P	<b>0.9988</b>	1.0152	<b>0.1068</b>	<b>0.0717</b>	<b>2.0476</b>

### Assessment of wind power and energy:

To estimate the turbine productivity, it is necessary to calculate wind power and AEP to be obtained from each fitted model. If the density of the air  $\rho = 1.16 \text{ kg/m}^3$  and the turbine rotor diameter  $D = 32 \text{ m}$ , the expected annual wind power and energy achieved from each site is determined using equations (22-25) for both 2-p and 3-p Weibull distributions. Table 8 shows estimated results for wind power and AEP.

The relative error of the estimated power shown in Table 18 is calculated as follows:

$$\text{Relative error (\%), RE} = \frac{\text{Estimated power} - \text{Actual power}}{\text{Actual power}} \times 100\%$$

If the RE is close to zero, the method estimates the parameter accurately. However, a positive RE implies an over-estimate and a negative RE implies an under-estimate by the method. From the values of Power, RE and AEP presented in Table 8, we observed that the BAYESIAN.2P and BAYESIAN.3P are found to be most efficient methods for estimating power and AEP. The BAYESIAN.3P procures the most accurate wind power and AEP for the sites 1, 2, 3, 5 and 7, which is very close to the actual power of these sites with smaller RE compared to other methods. On her other hand, the BAYESIAN.2P provides the best estimate of power and AEP for the sites 4 and 6 with smaller relative error.

Table 8: Estimated power, relative error in power estimate and energy production (Wind speed: 0 m/s – infinity)

Site	Methods	Power (kW)	RE (%)	AEP (kWh)
1	Actual	108457	0.00	950083
	MLE.2P	110306	1.70	966279
	MLE.3P	110061	1.47	964134
	BAYESIAN.2P	110298	1.70	966206
	BAYESIAN.3P	<b>110051</b>	<b>1.47</b>	<b>964047</b>
2	Actual	133525	0.00	1169683
	MLE.2P	136039	1.88	1191705
	MLE.3P	135765	1.68	1189301

	BAYESIAN.2P	136030	1.88	1191623
	BAYESIAN.3P	<b>135753</b>	<b>1.67</b>	<b>1189195</b>
3	Actual	206184	0.00	1806168
	MLE.2P	220464	6.93	1931263
	MLE.3P	209474	1.60	1834996
	BAYESIAN.2P	220556	6.97	1932074
	BAYESIAN.3P	<b>209458</b>	<b>1.59</b>	<b>1834856</b>
4	Actual	322492	0.00	2825032
	MLE.2P	319732	-0.86	2800855
	MLE.3P	319627	-0.89	2799928
	BAYESIAN.2P	<b>319812</b>	<b>-0.83</b>	<b>2801557</b>
	BAYESIAN.3P	319591	-0.90	2799619
5	Actual	290717	0.00	2546683
	MLE.2P	291809	0.38	2556248
	MLE.3P	290928	0.07	2548528
	BAYESIAN.2P	291696	0.34	2555257
	BAYESIAN.3P	<b>290889</b>	<b>0.07</b>	<b>2548188</b>
6	Actual	249192	0.00	2182919
	MLE.2P	247792	-0.56	2170656
	MLE.3P	247770	-0.57	2170468
	BAYESIAN.2P	<b>247848</b>	<b>-0.54</b>	<b>2171151</b>
	BAYESIAN.3P	247808	-0.56	2170799
7	Actual	204938	0.00	1795257
	MLE.2P	207543	1.27	1818077
	MLE.3P	204781	-0.08	1793882
	BAYESIAN.2P	207487	1.24	1817586
	BAYESIAN.3P	<b>204790</b>	<b>-0.07</b>	<b>1793960</b>

Thus, the comparison results, based on goodness of fit and the power estimation at different sites in the Equatorial region, show that the **3-p Weibull with Bayesian estimates the method to be used for wind energy resource assessments**. If any sites have lower occurrences of the lower wind speeds, then the shift parameter  $\theta$  in the 3-p Weibull distribution will be close to or equal to zero, which will be the 2-p Weibull distribution with Bayesian estimates. Another advantage of Bayesian approach is that it will reduce the need for long-term measurements for assessing the wind power potential of a site. This is possible with the integration of prior information with short-term data from a candidate site or historical data from one neighboring survey station.

## 9 Conclusion

Knowledge of correct statistical distribution of wind speeds at a given site is very important for accurate wind resource assessment. Some sites provide high uncertainty while fitting the traditional two-parameter Weibull distributions to wind speed data and warrant the need to explore distributions that characterize wind speeds better, such as the three-parameter Weibull distribution which is also a generalized form of two-parameter Weibull distribution with an additional non-zero shift parameter. In this study, investigation of wind characteristics and wind energy potential are carried out at different locations in the Equatorial region, of which three sites are in Fiji and one each from Cooks Islands, Tonga, Kiribati and Vanuatu, respectively. The wind speed data at these seven sites were tested for the best model between the two-parameter and three-parameter Weibull distributions. Furthermore, as there is no unique method that characterizes wind data perfectly, it is also imperative to have the knowledge of the best method of estimation for the parameters of wind speed distribution at a given site. In this study, we also introduced a novel

approach by using the Bayesian method for estimating parameters of wind speed distributions at the seven sites selected for testing the method. Then, a comparison study was conducted for the robust performance of the proposed Bayesian method with the popular frequentist MLE method. Finally, the results suggest that the three-parameter Weibull distributions should be used in analyzing wind power potential irrespective of a location. When the wind distribution has frequent low wind speeds and is less skewed, a three-parameter Weibull distribution is found to be a better fit. On the other hand, when wind distribution has less frequent low speeds and highly skewed, the two-parameter Weibull distribution which is a special case of three-parameter distribution with zero shift parameter is found to be more appropriate. The results also indicate that the Bayesian approach provides more accurate results while characterizing wind speeds and can be proposed as an alternative technique for estimating Weibull parameters. The proposed method can be incorporated in the popular software packages such as WAsP used for wind resource assessment and for planning wind energy projects.

### Data accessibility statement

Sample data are provided with the manuscript. Full data will be made available upon request and after approval from the respective Governments.

### References:

- Al Omari, M. A. and Ibrahim, N. A. (2011) Bayesian survival estimation for Weibull distribution with censored data. *Journal of Applied Sciences*, 11 (2):393-396.
- Azad, A.K., Rasul, M.G. and Yusaf, T. (2014) Statistical diagnosis of the best Weibull methods for wind power assessment for agricultural applications. *Energies*, 7:3056–3085. DOI:10.3390/en7053056.
- Aukitino, T., Khan, M.G.M., Ahmed, M.R. (2017) Wind energy resource assessment for Kiribati with a comparison of different methods of determining Weibull parameters. *Energy Conversion and Management*, 151:641–660.
- Carta, J.A., Ramírez, P. and Velázquez, S. (2009) A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands. *Renewable and Sustainable Energy Reviews*, 13(5):933-955.
- Casella, G. and Berger, R.L. (2002) *Statistical inference*. 2nd ed. USA: Thomson Learning Inc.
- Chaurasiya, P.K., Ahmed, S. and Warudkar, V. (2018) Comparative analysis of Weibull parameters for wind data measured from met-mast and remote sensing techniques. *Renewable Energy*, 115(C): 1153-1165.
- Chang, T.P. (2011) Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application. *Applied Energy*, 88(1):272-282.
- Corotis, R.B., Sigl, A.B. and Klein, J. (1978) Probability models of wind velocity magnitude and persistence. *Solar energy*, 20(6):483-493.
- Cousineau, D. and Allan, T. A. (2015) Likelihood and its use in parameter estimation and model comparison. *Mesure et évaluation en éducation*, 37(3):63-98. DOI:10.7202/1036328ar.
- Gosai, A.D. (2014) Wind Energy potential and electricity generation at Rokavukavu, Rakiraki. *MSc thesis*, The University of the South Pacific.
- Guure, C.B., Ibrahim, N.A. and Al Omari, M.A. (2012) Bayesian estimation of two-parameter Weibull distribution using extension of Jeffreys' prior information with three loss functions. *Mathematical Problems in Engineering*, 12. DOI:10.1155/2012/589640.

- Hossain, A. and Zimmer, W. (2003) Comparison of estimation methods for Weibull parameters: complete and censored samples. *Journal of Statistical Computation and Simulation*, 73 (2): 145–153.
- IEC (2017) Wind turbines – part 12-1: power performance measurements of electricity producing wind turbines. Available at: [www.iec.ch](http://www.iec.ch) (accessed 25 November 2020).
- Indhumathy, D., Seshaiyah, C.V. and Sukkiramathi, K. (2014) Estimation of Weibull parameters for wind speed calculation at kanyakumari in India. *International Journal of Innovative Research in Science*, 3(1):8340-8345.
- Jain, P. (2016) *Wind Energy Engineering*. New York, NY: McGraw Hill.
- Kidmo, D.K., Danwe, R., Doka, S.Y. and Djongyang, N. (2015) Statistical analysis of wind speed distribution based on six weibull methods for wind power evaluation in Garoua, Cameroon. *Revue des Energies Renouvelables*, 18(1):105–125.
- Lawless J.F. (2003) *Statistical Models and Methods for lifetime Data*, second ed. John Wiley & Sons, Hoboken, New Jersey.
- Luceño, A. (2008) Maximum likelihood vs. maximum goodness of fit estimation of the three-parameter Weibull distribution, *Journal of Statistical Computation and Simulation*, 78(10): 941-949. DOI: 10.1080/00949650701467363.
- Lye, L.M., Hapuarachchi, K.P. and Ryan, S. (1993) Bayes estimation of the extreme-value reliability function. *IEEE Transactions on Reliability*, 42(4):641-644. DOI: 10.1109/24.273598.
- Mohanty, M. (2012). New renewable energy sources, green energy development and climate change: Implications to Pacific Island countries. *Management of Environmental Quality: An International Journal*, 23(3):264-274. DOI:10.1108/14777831211217468.
- Mostafaiepour, A., Jadidi, M., Mohammadi, K. and Sedaghat, A. (2014) An analysis of wind energy potential and economic evaluation in Zahedan, Iran. *Renewable and Sustainable Energy Reviews*, 30:641-650.
- Manwell, J.F., McGowan, J.G. and Rogers, A.L. (2010) *Wind energy explained: theory, design and application*. Wiley.
- Pacific Power Association (2015) World Bank Provides Grant for Renewable Energy, available online at <http://www.ppa.org.fj/world-bank-provides-grant-for-renewable-energy/>
- Pandey, B.N., Dwividi, N. and Pulastya, B. (2011) Comparison between Bayesian and maximum likelihood estimation of the scale parameter in Weibull distribution with known shape under linex loss function. *Journal of Scientific Research*, 55:163–172.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3<sup>rd</sup> International Workshop on Distributed Statistical Computing (DSC 2003)*, March 20-22, Vienna, Austria, ISSN 1609-395X.
- Ramachandran, K.M. and Tsokos, C.P. (2015) *Mathematical statistics with applications in R*. Elsevier.
- Spiegelhalter, D.J., Best, N., Carlin, B.P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of Royal Statistical Society, Series B*, 64:583-639.
- Su, Y. and Yajima, M. (2020). *Using R to run 'JAGS'*. R package version 0.6-1, URL <https://cran.r-project.org/web/packages/R2jags>.
- Sukkiramathi, K. and Seshaiyah, C.V. (2020) Analysis of wind power potential by the three-parameter Weibull distribution to install a wind turbine, *Energy Exploration & Exploitation*, 38(1):158-174. DOI: 10.1177/0144598719871628

- Teimouri, M. and Gupta, A.K. (2013). On the three-parameter Weibull distribution shape parameter estimation. *Journal of Data Science*, 11:403-414.
- Tuller, S.E. and Brett, A.C. (1984). The Characteristics of Wind Velocity that Favor the Fitting of a Weibull Distribution in Wind Speed Analysis, *Journal of Climate and Applied Meteorology*, 23(1), 124-134.
- VNR Report (2019), Voluntary National Review, Republic of Fiji, 38-41
- Wais, P. (2017) Two and three-parameter Weibull distribution in available wind power analysis. *Renewable energy*, 103: 15-29.

### Author contributions

MGMK: Conceptualization of new method, writing the code, simulations and draft paper; MRA: Data acquisition, compilation, validation and review of manuscript.

### Declaration of competing interests

The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

Funds for carrying out this work were provided by Korea International Cooperation Agency (KOICA) under its East-Asia Climate Partnership program. The project number was 2009-00042.

### ORCID ID

M.G.M. Khan: <https://orcid.org/0000-0001-5400-9703> and M. Rafiuddin Ahmed: <https://orcid.org/0000-0002-3514-1327>

## Appendix 1

### Model specification for 2-parameter Weibull distribution

```
cat("
  model{
    # Likelihood
    for (i in 1:length(U)){
      p[i] <- dweib(U[i],shape, lambda);
      observed[i] ~ dbern(p[i]);
    }

    # Priors
    shape ~ dunif(0,4) # After a series of trial and error guesses
    scale ~ dunif(0,100)
    lambda <- pow(1/(scale), shape)
  }
  ", file="weibull_model_2p.txt")
```

### Calling JAGS from R

```
jags.fit <- jags(data, inits, parameters.to.save, n.iter=10000,
model.file="weibull_model_2p.txt",n.chains = 3, n.burnin = 1000,
n.thin=5)
```

## Model specifications for the 3-parameter Weibull distribution

```
cat("  
  model{  
    # Likelihood  
    for (i in 1:length(U)){  
      p[i] <- dweib(U[i] - shift ,shape, lambda) / 1000  
      observed[i] ~ dbern(p[i]);  
    }  
  
    # Priors  
    shape ~ dunif(0,4)  
    scale ~ dunif(0,31)  
    shift ~ dunif(-1, 1)  
    lambda <- pow(1/(scale), shape)  
  }  
  ", file="weibull_model_3p.txt")
```

## Appendix 2

Trace and posterior density plots that show the convergence of Bayesian models for sites 1, 2, 5, 6 and 7.

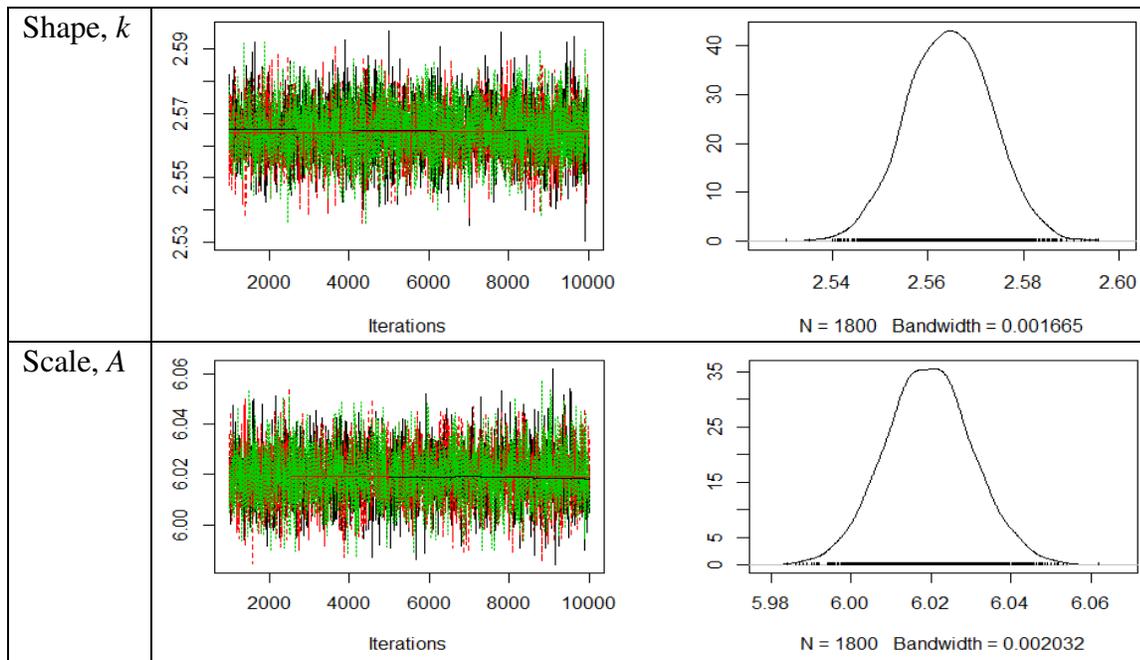


Figure 12: Trace and posterior density plots for site 1 (2-p Weibull).

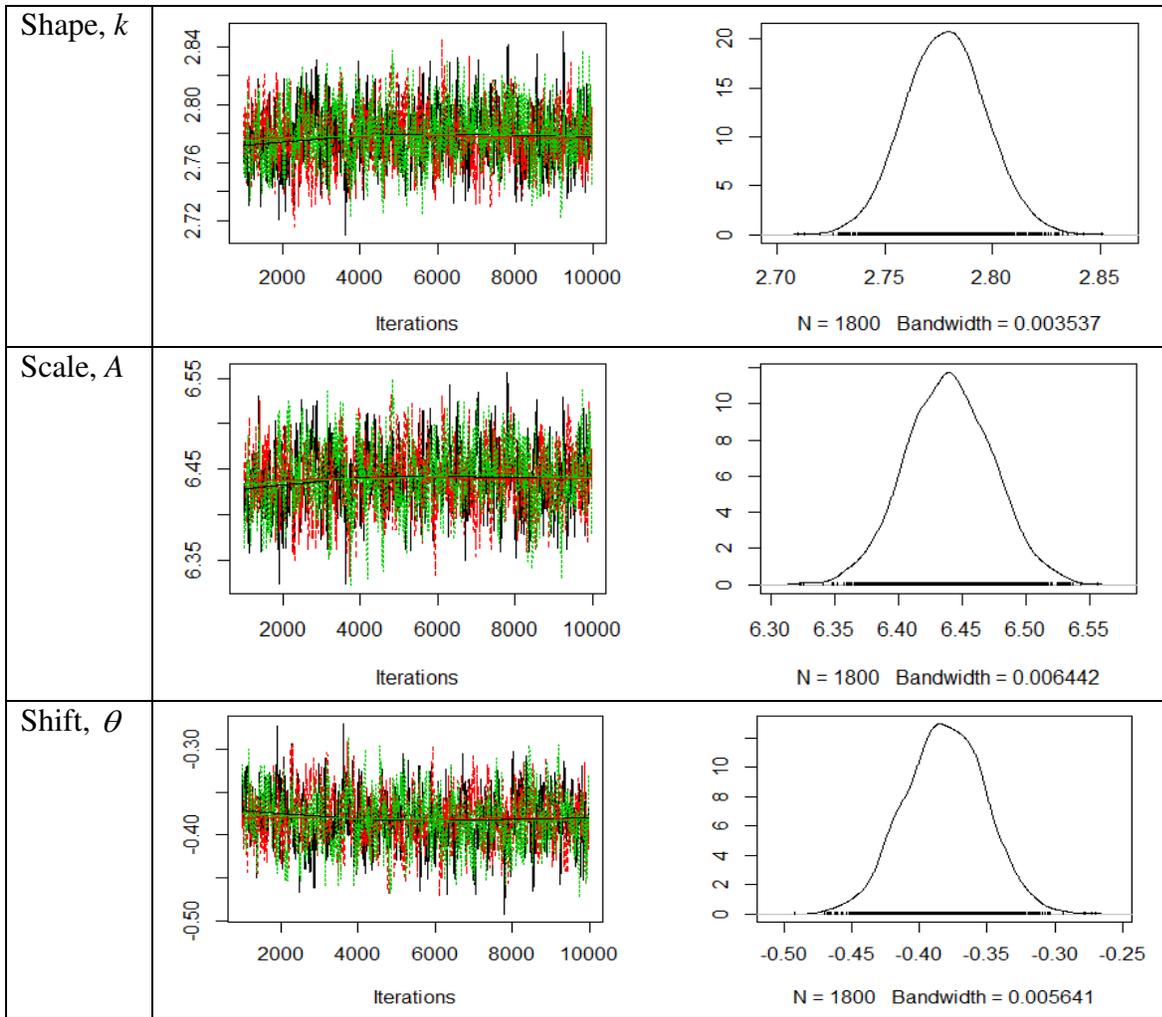
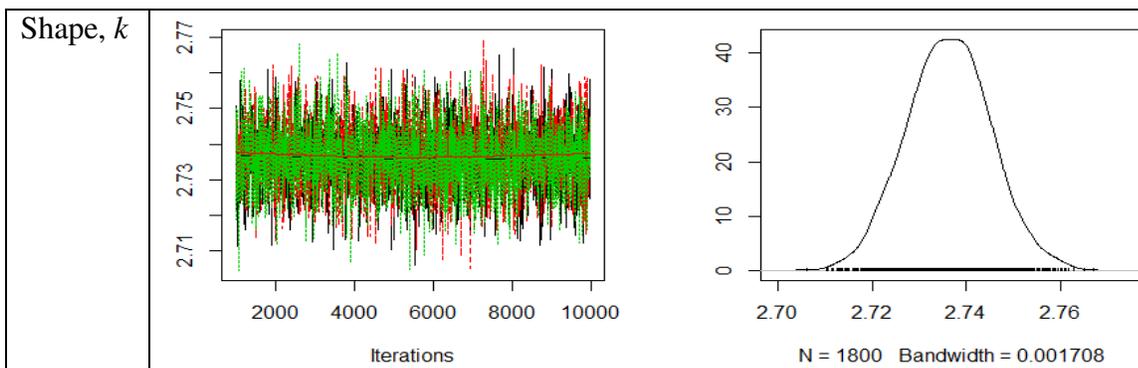


Figure 13: Trace and posterior density plots for site 1 (3-p Weibull).



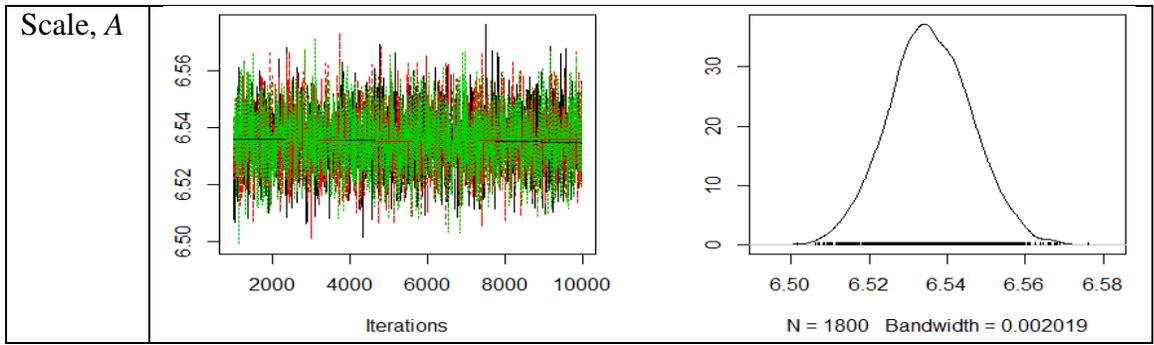


Figure 14: Trace and posterior density plots for site 2 (2-p Weibull).

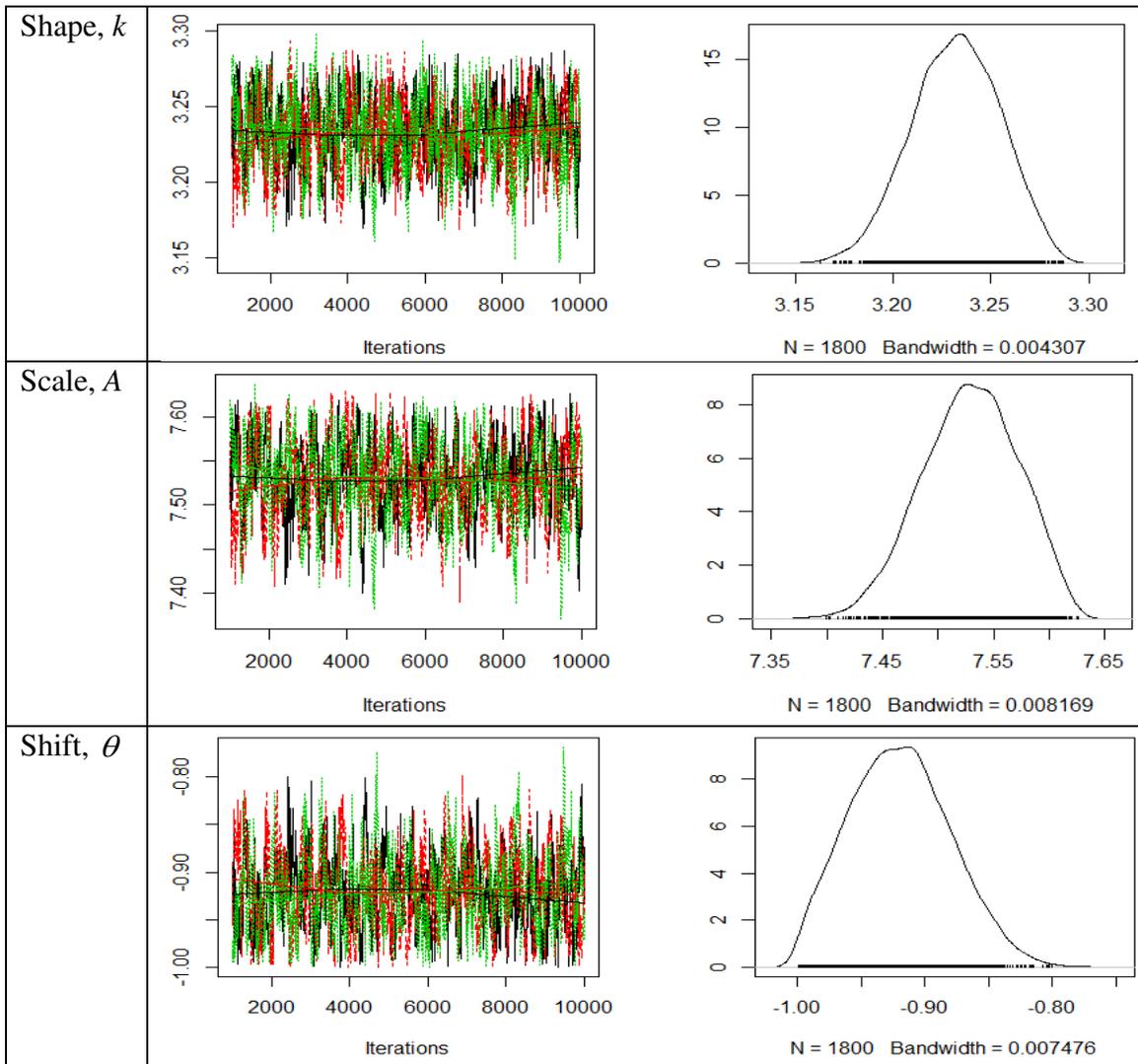


Figure 15: Trace and posterior density plots for site 2 (3-p Weibull).

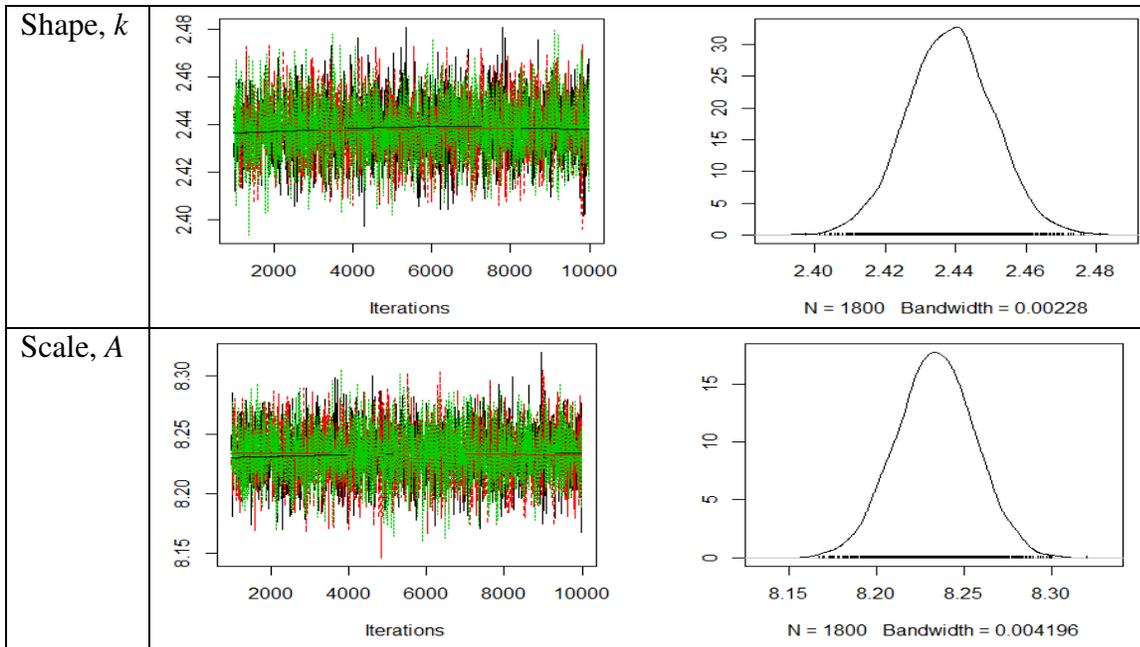


Figure 16: Trace and posterior density plots for site 5 (2-p Weibull).

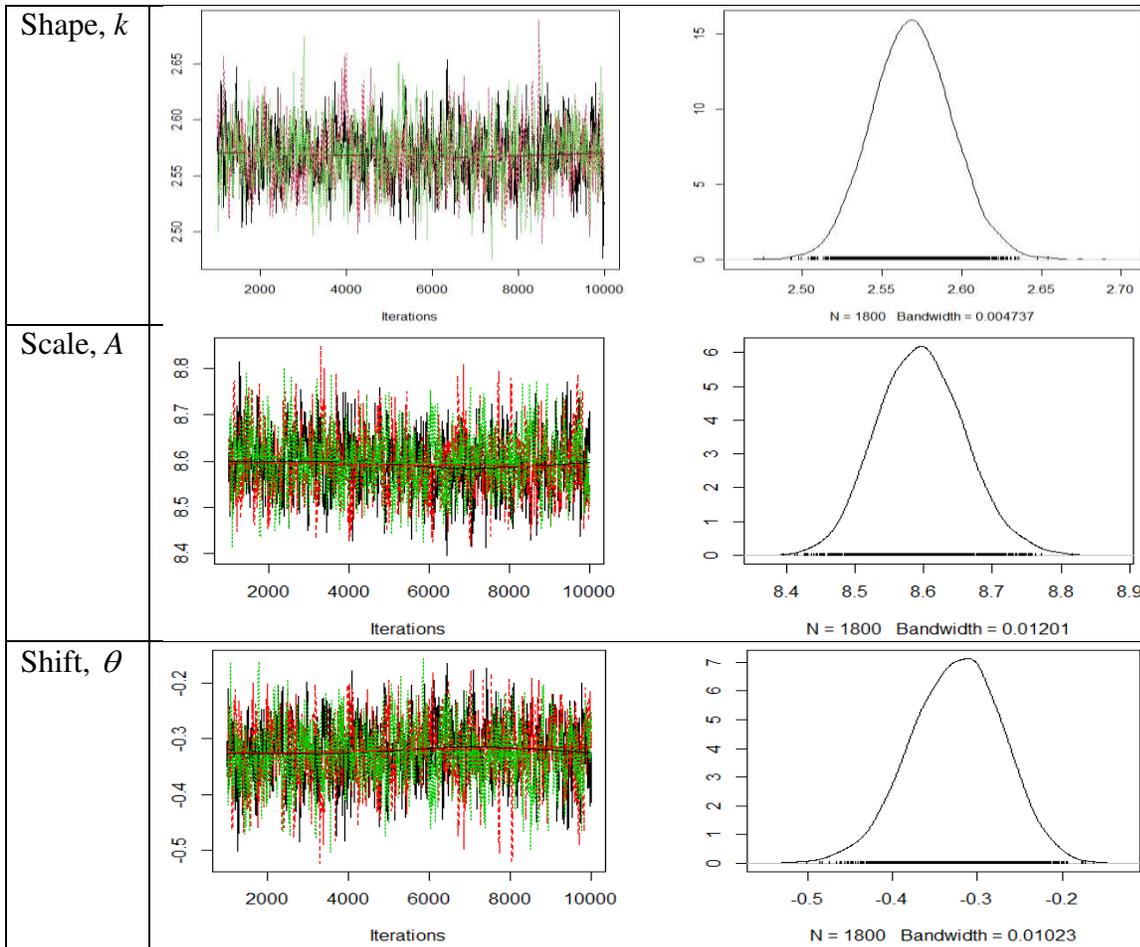


Figure 17: Trace and posterior density plots for site 5 (3-p Weibull).

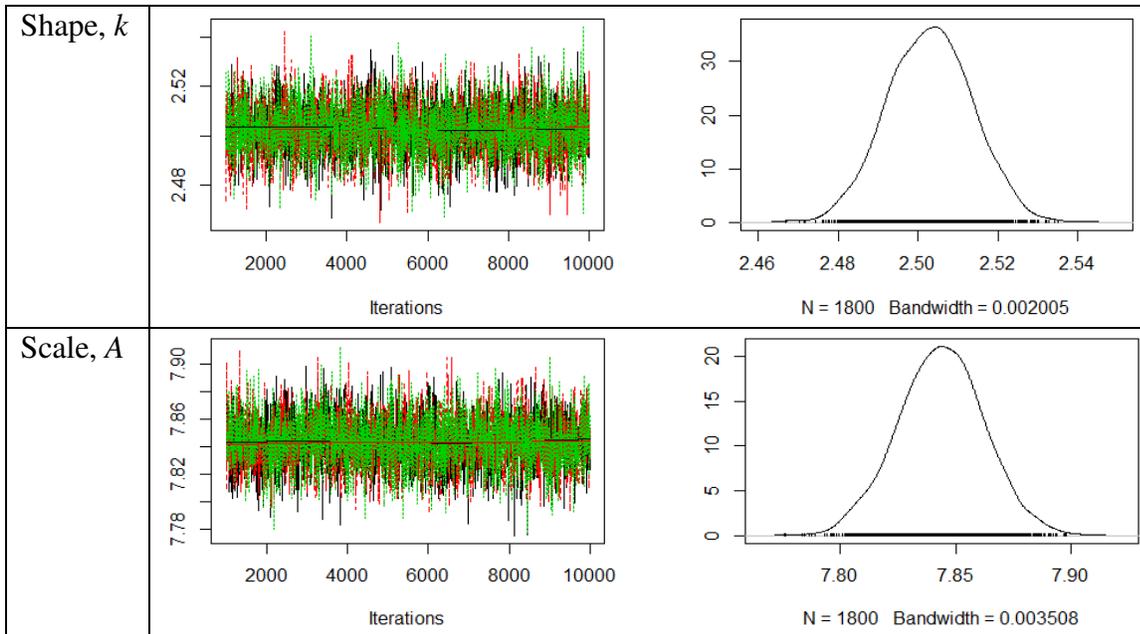


Figure 18: Trace and posterior density plots for site 6 (2-p Weibull).

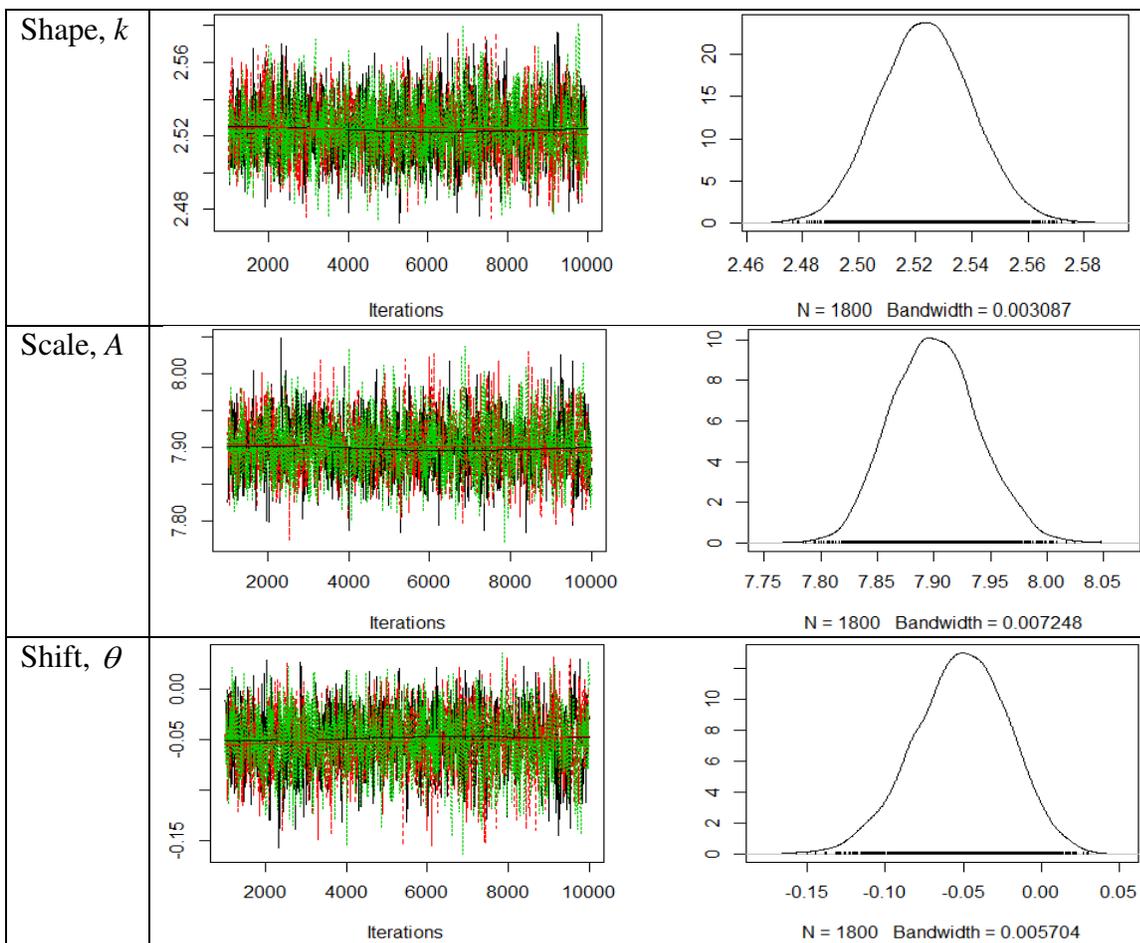


Figure 19: Trace and posterior density plots for site 6 (3-p Weibull).

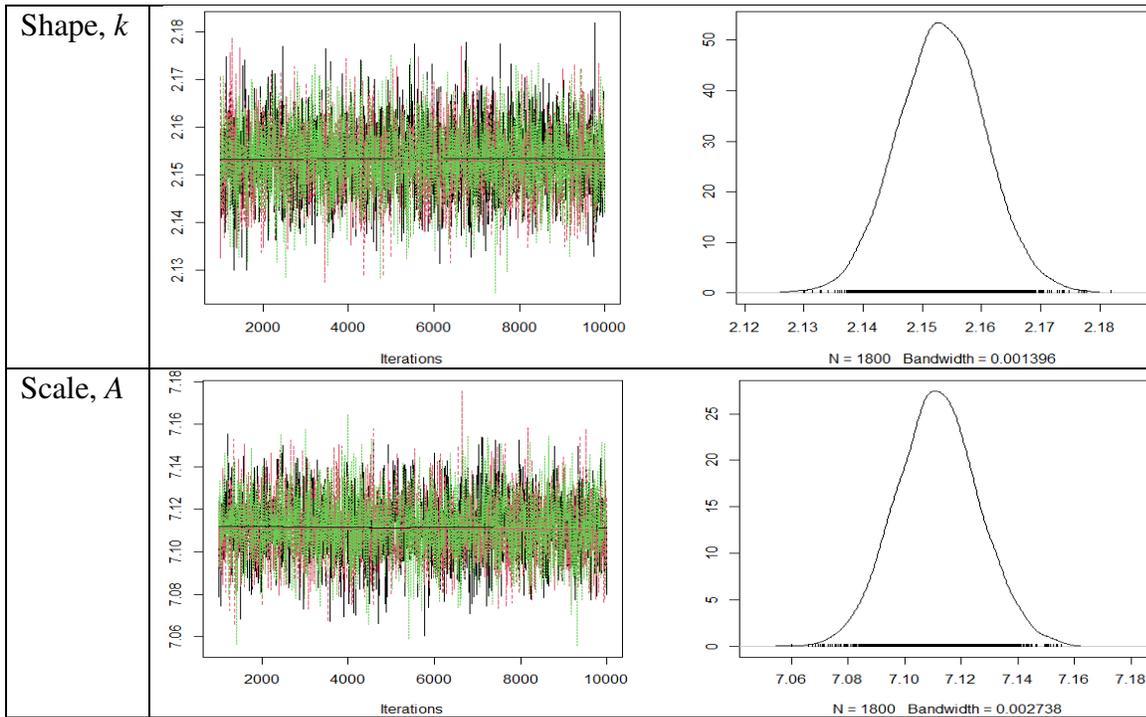
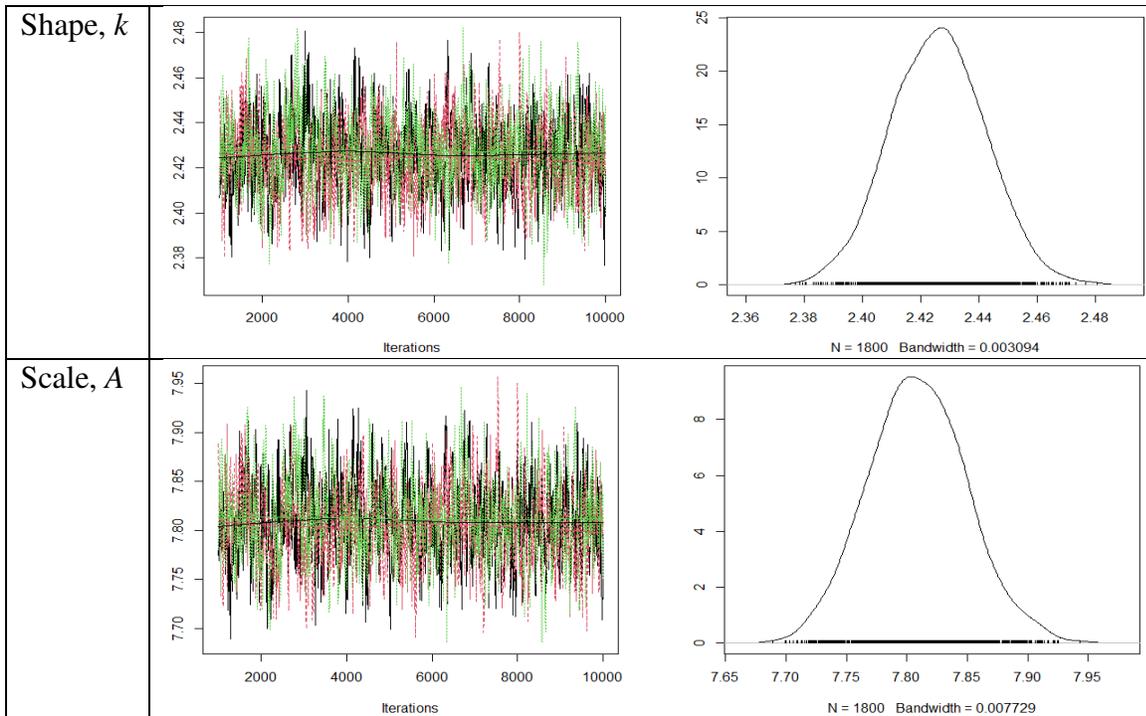


Figure 20: Trace and posterior density plots for site 7 (2-p Weibull).



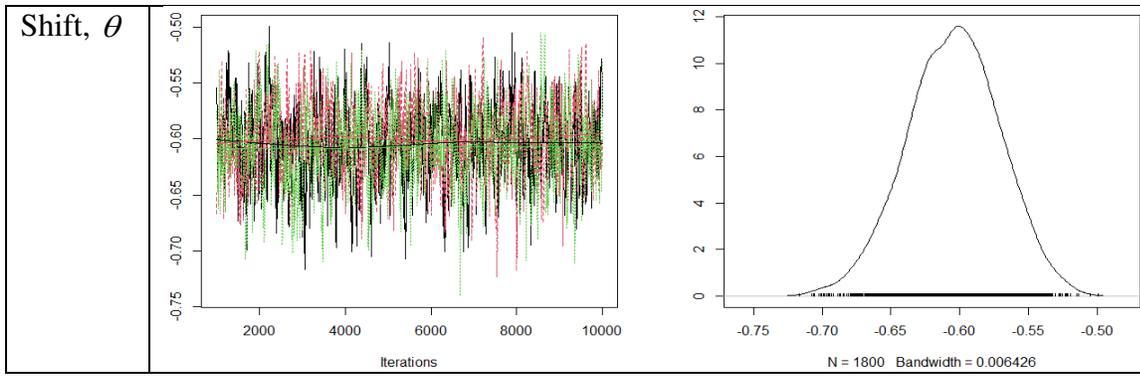


Figure 21: Trace and posterior density plots for site 7 (3-p Weibull).

**Declaration of interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

M. G. M. Khan