Author Query Form

Journal RPG2

Article rpg212297

Dear Author,

During the copyediting of your manuscript the following queries arose.

Please refer to the query reference callout numbers in the page proofs and respond.

Please remember illegible or unclear comments and corrections may delay publication.

Many thanks for your assistance.

Query No.	Description	Remarks
Q1	Please confirm that forenames/given names (blue) and surnames/family names (vermilion) have been identified correctly.	
Q2	Please provide full name of all the authors.	
Q3	Please check whether these asterisks should be replace with some variable.	
Q4	Please expand the term CDF.	
Q5	Please check the inserted Conference location and dates in ref 14.	

Received: 20 April 2021 | Revised: 9 July 2021 | Accepted: 26 August 2021

IET Renewable Power Generation

DOI: 10.1049/rpg2.12297

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

Bayesian estimation of copula parameters for wind speed models of dependence

Saul B. Henderson[1] | Amir Shahirinia[1] | M. Tavakoli Bina[2]

[1] Department of Electrical & Computer Engineering, University of the District of Columbia, Washington, DC, USA

[2] Department of Electrical Engineering, K.N. Toosi University of Technology, Tehran, Iran

Correspondence
M. Tavakoli Bina, Department of Electrical Engineering, K.N. Toosi University of Technology, P.O Box 16315-1355, Tehran, Iran.
Email: tavakoli@eetd.kntu.ac.ir

Abstract

Modelling the uncertainty of wind speed is essential in power flow analysis. Having abundant knowledge of the wind speed in an area is critical. A low volume of data can increase uncertainty in wind speed analysis. Spatial dependencies are often modelled before running probabilistic power flow and load flow analysis. Copulas are a popular way of capturing spatial dependence between multiple wind farms. Using NREL data from seven Northeastern United States wind farm sites, Bayesian inference will be used to determine the copula parameter uncertainty between weekly, daily, and hourly wind speed observations. This approach will be used on elliptical and single parameter Archimedean copulas. For each possible wind farm pair, an uninformative prior will be placed on the copula parameter. The resulting posterior will contain a distribution of copula parameter values based on the prior and the observed wind speed data. The posterior's credible interval is reviewed to determine the uncertainty in parameter estimation. The results show that using a data volume considerably more petite than 8760 hourly data points will result in more uncertainty in parameter estimation and inaccuracies in wind speed forecasting if using non-Bayesian methods for copula parameter estimation.

1 | INTRODUCTION

Studying the random behaviour of wind energy sources is of great importance as it influences cascading failures within a power system [1] and adversely affects reliability and security [2]. Various studies of power analysis concerning the random behaviour of wind speed have been published, specifically using probabilistic models such as probabilistic power flow, optimal power flow, and power load flow [3–6]. More probabilistic models are often required to capture the various dependencies among wind farms prior to probabilistic flow analysis of wind resources. Copulas can reliably model pair-wise spatial dependence [7, 8], multiple spatial dependence using vine copulas [9], and temporal correlation [10] of the wind farms prior to power flow and load flow analysis.

While higher dimensional copulas can be hard to simulate or computationally demanding [10, 11], pairwise copulas can be reliably modelled and manipulated in many coding environments such as MATLAB, Python, and R. Because there is a fair amount of support for pairwise copulas, one can easily build and simulate bivariate distributions to simulate wind speed data points for power flow analysis. Distributions using pairwise copulas require the proper marginal distributions to fit the wind speed data. In literature, the Weibull distribution is commonly accepted as the distribution to model wind speed data [12–14], while some use the lognormal and gamma distributions [11, 15]. For copula modelling, the best copula and its parameter for dependence modelling is often selected based on maximum likelihood estimation, the Akaike information criterion (AIC) and/or Bayesian information criterion (BIC). As a result, a singular copula parameter is expected to capture dependence among two wind farms.

A smaller volume of data will bring far more uncertainty in choosing the parameter than using large data sets because smaller data sets do not provide enough observations to form a distinctive dependence structure. To capture the uncertainty of parameter selection, distribution of likely parameters is required instead of a single value. This distribution will depend on the wind speed data available and some knowledge about the copula parameter in question. The best method that could provide such a distribution is Bayesian inference. Using some knowledge about the parameter (prior) and the available wind speed data (likelihood), a distribution of parameters from the posterior can be obtained. To capture the uncertainty, the credible interval (CI) of the posterior is examined to assess the deviation of values from the mean.

Bayesian inference returns a distribution of values to analyse. Furthermore, the certainty of the returned parameters can be assessed based on the deviation from the posterior mean, and the certainty in copula parameter selection can be assessed for data sets with varying number of observations. In this paper, Bayesian inference will be used to analyse the effect of wind speed data volume on copula parameter estimation for elliptical copulas and single-parameter Archimedean copulas. Wind speed data from several wind farm sites within the Northeastern Unites States will be used in this study. The uncertainty of each copula parameter calculated for each wind farm pair from the posterior's CI will be examined. The rationale of using Bayesian inference instead of likelihood estimation is threefold:

1. The Bayesian inference process will return the possible choices of copula parameters for a certain wind farm pair that can all be examined. This is advantageous for smaller data sets with a low observation count.
2. Using weekly, daily and hourly observations, the Bayesian process will confirm the effect of data set size on the copula parameter estimation process.
3. Bayesian inference will determine what data set size is suitable for copula parameter point-estimation using non-Bayesian techniques (i.e. maximum likelihood, AIC, BIC).

The rest of this paper is organized as follows: Section 2 will provide background on the copulas used in this study and Bayesian inference. Section 3 will provide an overview of the data used. Section 4 will provide the simulation and results. We conclude our work in Section 5.

2 | BACKGROUND

2.1 | Bayesian inference

To model uncertainty in our estimation of the copula parameter, we turn to Bayesian inference: the process of updating our beliefs with the data presented.

There are three elements involved in Bayesian inference: prior, likelihood and the posterior. The prior is a distribution that represents what we know about the parameter. Depending on the distribution chosen as the prior, our knowledge of the parameter can be accurately represented. The likelihood is an equation that is used to model the data with each value from the prior distribution. In our case, the likelihood equation will be the copula density function of the intended copula. The result of the prior and likelihood is the posterior – which reflects our updated beliefs given the data. From the posterior, the CI is examined for the deviation from the posterior mean. Bayesian inference is based on the Bayes Theorem:

$$\pi(\theta|x) = \frac{\pi(\theta)\,L(x,\theta)}{\int \pi(\theta)\,L(x,\theta)} \equiv \pi(\theta)\,L(x,\theta) \tag{1}$$

where $\pi(\theta|x)$ represents the posterior and the right side is the prior times the likelihood. While the equation in the middle is the true formula for Bayesian inference, the denominator (marginal data) is often left out as it is hard to estimate. Markov Chain Monte Carlo (MCMC) algorithms are used to sample from the posterior when a closed-form equation is unavailable. Simply multiplying the prior and the likelihood may not give exact results for the posterior, so many iterations must be performed for MCMC for a large distribution of values to be returned.

Using Equation (1), we can mathematically model our Bayesian approach for copula parameter estimation. To build the model, a suitable prior and likelihood is required. An informative or an uninformative prior can be chosen to represent our knowledge of the parameters. Since there is little knowledge of the parameter, a uniform (flat) prior, $f_{unif}(\theta; a, b)$, will be used to represent an uninformative prior. For the likelihood density function, the copula density function ***** will be used. Q3
Since the copula density function is a joint probability distribution function of uniform margins u and v, this function can be used to determine which sampled parameter best models the data u and v. The general formula for Bayesian copula parameter estimation will be:

$$\pi(\theta|u,v) = f_{unif}\ (\theta; a, b) * \sum_{i=1}^{N} \log(c(u_i, v_i|\theta)) \tag{2}$$

where $\pi(\theta|u,v)$ is the desired posterior, N is the number of observations in the data, and $f_{unif}(\theta, a, b)$ is the uniform prior with a and b boundaries. The copula density function can vary based on the copula used. For certain copulas, the association parameter has bounds that must be adhered to. For example, the normal and t-copulas use Spearman's rho as the association parameter. Rho has values between 0 and 1 so the prior's bounds must adhere to that. For Archimedean copulas, the association parameter can be infinite. For this study, the uniform prior will be used.

2.2 | Bayesian approach to copula modelling

The copula is based on Sklar's Theorem [16] which states the following:

Let H be a joint distribution function with margins F and G. Then there exists a copula, C, such that for all x and y:

$$H(x,y) = C(F(x), G(y)) \tag{3}$$

As a linking function between marginal distributions and joint distributions, the copula captures linear correlation, tail dependence [17] and central dependence. Only three single-parameter copulas are used for Bayesian inference in this study: Normal, Student-t, and Frank. Other copulas were selected as the best copula for some wind farm pairs (BB1, BB8), but Bayesian inference was not performed. These copulas are defined in the Appendix.

Two important consequences of a Bayesian predictive copula model are as follows:

a. Inclusion of uncertainty about the parameters of the dependent wind speed distribution (copula) results in using a more prudent predictive multivariant distribution for the correlated wind speeds. That means, on average, the Bayesian predictive copula probability distribution is more dispersed than the non-Bayesian copula probability distributions when the uncertainty about the parameters is ignored. Consequently, for example, for a range of the wind speed with a given probability, the range under the Bayesian predictive copula distribution is wider than that of ignoring the parameter uncertainty. Conversely, the wind speed range under the Bayesian predictive copula distribution is narrower than that of neglecting the parameter uncertainty for a given probability.
b. The probability distributions of the copula parameters can be viewed in terms of the heterogeneity of the dependent wind speed distribution over the wind farms. The dependent wind speed distributions for various locations (farms) may belong to the same family of models (copula), such as normal, t, Frank, BB1, and BB8 copulas, but the model parameters may vary randomly according to some probability distributions (prior distributions) instead of being fixed values. Hence, the Bayesian predictive copula distribution aggregates the non-homogeneous copula distributions into a single copula distribution that captures the variation among the probability distributions of the dependent wind speeds at the farms' locations.

2.3 | Elliptical copulas

The normal (Gaussian) and Student-t copulas are known as the elliptical copulas due to their symmetrical nature [18].

2.3.1 | Normal copula

The normal copula utilizes the bivariate normal distribution (and the linear correlation coefficient ρ) on u and v after both variables have been inversely transformed using the standard normal inverse CDF. So the equation will be:

$$C_\rho(u,v) = \Phi_\rho\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) \tag{4}$$

Denoting s as $\Phi^{-1}(u)$ and t as $\Phi^{-1}(v)$, the copula density function that will be used as the likelihood is defined as:

$$c_\rho(u,v) = \left(\frac{1}{\sqrt{1-\rho^2}}\right)\exp\left(\frac{-\rho^2 s^2 + \rho^2 t^2 - 2pst}{2(1-\rho^2)}\right) \tag{5}$$

Using the normal copula density and the uniform prior, the equation to estimate the normal copula parameter will be:

$$\pi(\rho|u,v) = f_{unif}(\rho; a, b)\sum_{i=1}^{N}\log\left(\frac{1}{\sqrt{1-\rho^2}}\right) * \exp\left(\frac{-\rho^2(s_i)^2 + \rho^2(t_i)^2 - 2\rho s_i t_i}{2(1-\rho^2)}\right) \tag{6}$$

2.3.2 | Student-t copula

Much like the normal copula, the Student-t copula utilizes transformed u and v variables to form its structure. Instead of standard normal transformations, u and v will be inversely transformed using the t-distribution, with η degrees of freedom, and will then be used as arguments in a bivariate t-distribution. The equation for the t-copula will be:

$$C_{\eta\rho}(u,v) = T_{\eta\rho}\left(T_\eta^{-1}(u), T_\eta^{-1}(v)\right) \tag{7}$$

The copula density function is defined as:

$$c_{\eta\rho}(u,v) = \left(\frac{\Gamma\left(\frac{\eta+2}{2}\right)\Gamma\left(\frac{\eta}{2}\right)}{\sqrt{1-\rho^2}\,\Gamma^2\left(\frac{\eta+1}{2}\right)}\right)\left(\frac{\left(1+\frac{s^2}{\eta}\right)\left(1+\frac{t^2}{\eta}\right)^{\frac{\eta+1}{2}}}{\left(1+\frac{s^2+t^2-2\rho st}{\eta(1-\rho^2)}\right)^{\frac{\eta+1}{2}}}\right) \tag{8}$$

where Γ is the gamma function $(n-1)!$. The Bayesian equation for estimating ρ will be:

$$\pi(\rho|u,v) = f_{unif}(\rho; a, b) * \sum_{i=1}^{N}\log\frac{\Gamma\left(\frac{\eta+2}{2}\right)\Gamma\left(\frac{\eta}{2}\right)}{\sqrt{1-\rho^2}\,\Gamma^2\left(\frac{\eta+1}{2}\right)} * \frac{\left(\left(1+\frac{(s_i)^2}{\eta}\right)\left(1+\frac{(t_i)^2}{\eta}\right)\right)^{\frac{\eta+1}{2}}}{\left(1+\frac{(s_i)^2+(t_i)^2-2\rho s_i t_i}{\eta(1-\rho^2)}\right)^{\frac{\eta+1}{2}}} \tag{9}$$

2.4 | Archimedean copulas

Unlike the elliptical copulas, Archimedean copulas utilize generator functions to work. Generator functions are strictly decreasing functions that map values from [0, 1] to values between [0,∞) [18]. Much like how distributions were used to transform u and v within elliptical copulas, generators have the same function within Archimedean copulas. In this study, the Frank copula is the only Archimedean copula used.

2.4.1 | Frank copula

The Frank copula is defined by the following generator and copula equations:

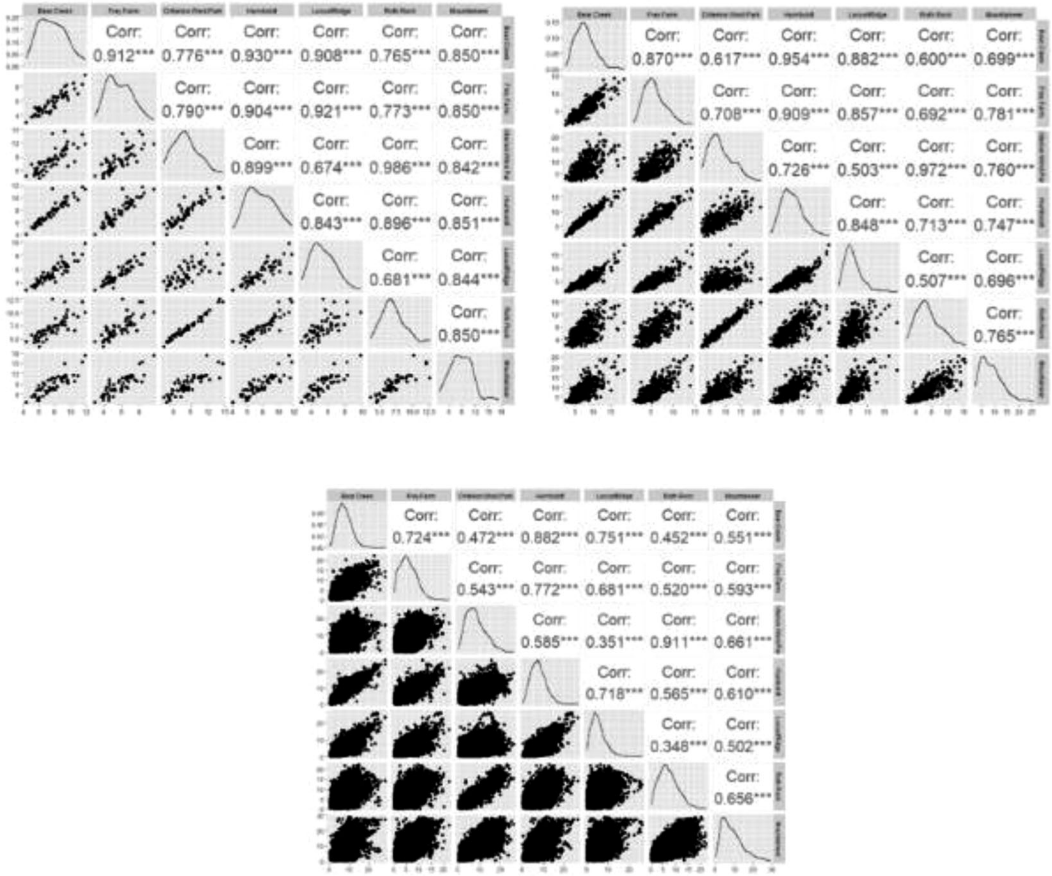

FIGURE 1 Scatterplots for wind speed under different volumes of data. In each scatterplot, the correlation between each wind farm is shown. Moreover, the marginal distribution for each wind farm is shown in the diagonal. For the weekly and daily wind speed data, the marginals can be represented by a lognormal or gamma distribution. For the hourly data set, the Weibull is used as the marginal distribution. In total, 21 possible wind farm pairs under each data set can be examined for spatial correlation. (Top Left) Weekly (53 points) wind speed data (Top Right) Daily (365) wind speed data (Bottom Centre) Hourly (8760 points) wind speed data

$$g(t) = -\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right) \tag{10}$$

$$C_\theta(u,v) = \frac{-1}{\theta}\ln\left(1+\frac{\left(e^{\theta u}-1\right)\left(e^{\theta v}-1\right)}{\left(e^{\theta}-1\right)}\right) \tag{11}$$

where the association parameter $\theta \neq 0$. The density function is defined as:

$$c_\theta(u,v) = \frac{\theta\left(1-e^{-\theta}\right)\left(e^{-\theta(u+v)}\right)}{\left(e^{-\theta}-1+\left(e^{-\theta(u)}-1\right)\left(e^{-\theta(v)}-1\right)^2\right)} \tag{12}$$

The Bayesian equation to estimate Frank copula parameters will be:

$$\pi(\theta|u,v) = f_{unif}\ (\theta; a, b) * \sum_{i=1}^{N} \log\frac{\theta\left(1-e^{-\theta}\right)\left(e^{-\theta(u_i+v_i)}\right)}{\left(e^{-\theta}-1+\left(e^{-\theta(u_i)}-1\right)\left(e^{-\theta(v_i)}-1\right)^2\right)} \tag{13}$$

Unlike the other copulas, the Frank copula does not model tail dependence but only central dependence.

3 | METHODOLOGY

Wind speed data from seven sites within Virginia, West Virginia, Maryland, and Pennsylvania have been gathered from the NREL Wind Prospector site for the year 2012. Using 8760-point wind speed data, a smaller data set is created by averaging values every 24 h to obtain a representation of daily (365-point) wind speed data. From the newly created daily data, we average values every 7 days to obtain a representation of weekly (53-point) wind speed data. These three data sets will be used in the uncertainty analysis of copula parameter estimation. Figure 1 shows the difference between each volume of data based on scatterplots.

Since both marginal distributions and a best-fitting copula are needed to build a multivariate distribution for each wind farm pair, a distribution must be first fitted to the wind speed distributions of each wind farm. This process must be done for all seven wind farms under all three data sets. This process can be done by first examining the skewness and kurtosis to get candidate

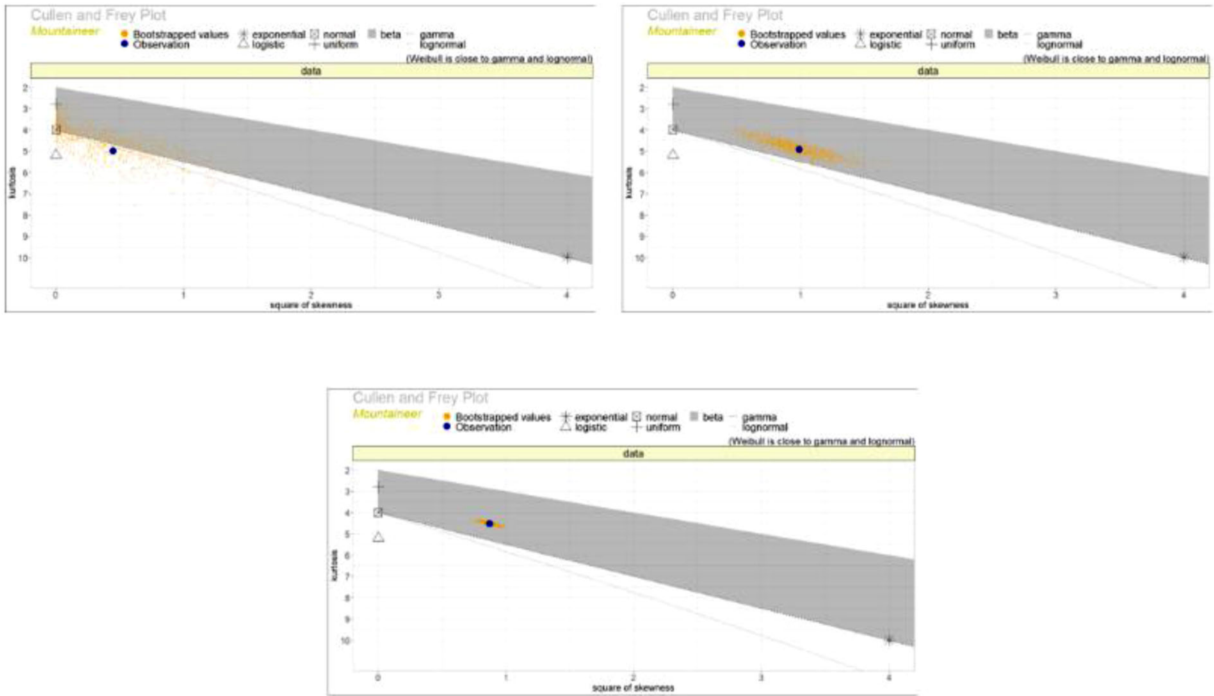

FIGURE 2 Cullen and Frey (Skewness vs Kurtosis) plots were used for determining the best-fitted distribution for Mountaineer – the seventh wind farm in this study. Under each data set, a reiterative process is done to create a thousand possible choices of distributions - represented by small orange dots. The large blue dot represents the optimal choice. As the volume of data increases, the certainty of distribution selection increases as shown in (c) (Top Left) weekly wind speed data for Mountaineer (Top Right) daily wind speed data for Mountaineer (Bottom Centre) hourly wind speed data for Mountaineer

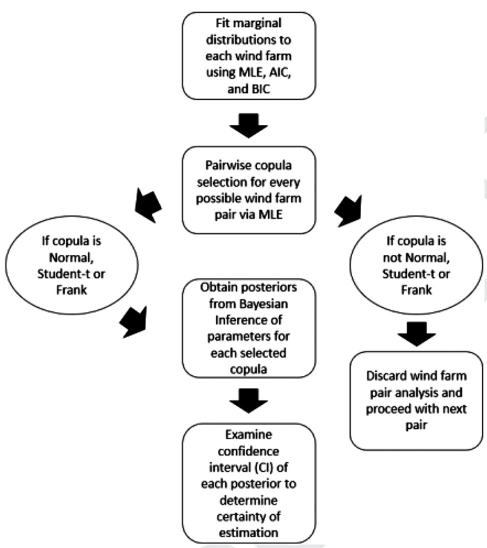

FIGURE 3 Flowchart of the complete methodology of the Bayesian approach for elliptical and single-parameter Archimedean copulas

TABLE 1 Selected wind farm locations

Wind farm number	Name	Location
1	Bear Creek	Bear Creek, PA
2	Frey Farm	Conestonga, PA
3	Criterion Wind Park	Garrett Co., MD
4	Humboldt	Hazleton, PA
5	Locust Ridge	Seltzer, PA
6	Roth Rock	Red House, MD
7	Mountaineer	Thomas, WV

distributions, and then using likelihood, AIC and BIC to choose the best distribution to fit the wind speed data. The Weibull distribution is commonly used to model wind speed distributions, while gamma or lognormal can model others. The Cullen–Frey plots in Figure 2 show a level of uncertainty in selecting a proper distribution for smaller data sets, but not for the 8760 hourly data points. The flowchart in Figure 3 depicts the complete methodology of the Bayesian approach for elliptical and single-parameter Archimedean copulas.

After determining proper distributions to fit the wind speed data, copula selection is performed for every possible wind farm pair. The relationship between copula selection, distance and wind farm correlation is examined. As each wind farm pair is within 480 kilometres apart from each other, the correlation is moderately high between each wind farm pair. The importance of spatial correlation has been studied in [19, 20], so such correlation will impact copula selection. It is expected that strong upper tail and central dependence will be captured. Copulas such as the normal, *t*, Frank, BB1 and BB8 may be used heavily as they can capture the dependence structure for such close wind farms.

TABLE 2 Bayesian copula estimation table

Distance (km)	52-point copula & confidence interval	52-point data posterior mean	365-point copula & confidence interval	365-point posterior mean	8760-point copula & confidence interval	8760-point posterior mean
48.2	Normal (0.84–0.91)	0.88	Student-*t* (0.85–0.89, *df* = 11)	0.868	Student-*t* (0.71–0.72, *df* = 13)	0.715
478.5	Frank (7.15–12.9)	9.89	Normal (0.58–0.63)	0.6	Student-*t* (0.43–0.46, *df* = 30)	0.45
490.5	Normal (0.79–0.88)	0.84	Normal (0.66–0.707)	0.685	Student-*t* (0.53–0.56, *df* = 30)	0.54
366	Normal (0.613–0.73)	0.673	Normal (0.51–0.562)	0.536	Normal (0.394–0.406)	0.397
40.2	Normal (0.972–0.988)	0.982	Student-*t* (0.97, *df* = 11)	0.97	Frank (13.97–14.56)	14.26
62.8	Normal (0.79–0.87)	0.84	Normal (0.74–0.78)	0.75	Normal (0.65–0.66)	0.655
446	Normal (0.79–0.873)	0.835	Normal (0.707–0.75)	0.72	Student-*t* (0.593–0.614, *df* = 18)	0.6
393	Frank (4.28–8.64)	6.4	Normal (0.51–0.56)	0.54	Normal (0.393–0.406)	0.397
23	Frank (8.18–14.43)	11.17	Normal (0.74–0.77)	0.75	Student-*t* (0.66–0.677, *df* = 18)	0.67

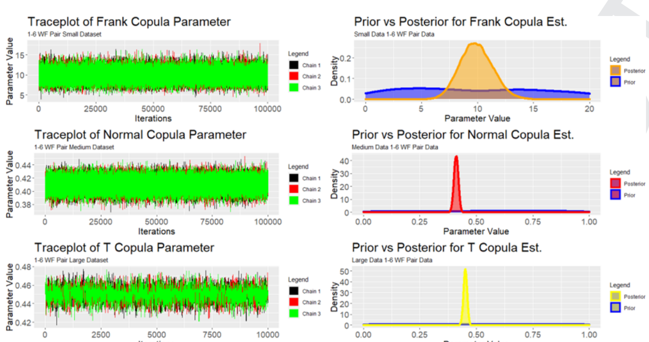

FIGURE 4 Trace plots of estimated copula parameters (left column) and prior versus posterior plots (right column) for the 1–6 (Bear Creek/Roth Rock) wind farm pair under each wind speed data set. The traceplots show the MCMC chains containing sampled copula parameters for each copula. The prior vs posterior plot shows the resulting posterior plotted against the prior (blue) to illustrate the confidence interval. While the Frank copula parameter can take values between 0 and ∞, the confidence integral is noticeably wide. The normal and *t*-copula parameters can only take values between 0 and 1. However, the confidence integral is small as shown in the posterior plots

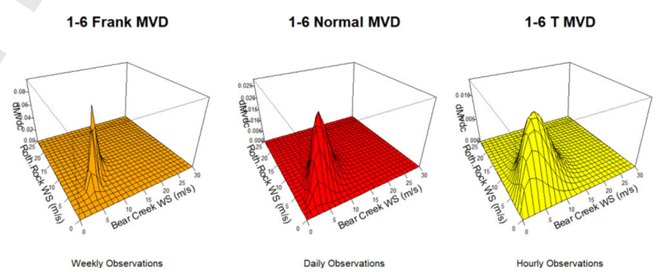

FIGURE 5 Multivariate distributions for the weekly (left), daily(middle) and hourly (right) of the 1–6 (Bear Creek/Roth Rock) wind farm pair using the chosen copulas for this pair. These distributions will be used to simulate weekly (Frank MVD), daily (normal MVD) and hourly (*t* MVD) wind speed data

Once the copulas have been selected, copula parameters for each wind farm pair under each data set are estimated. Only select single parameter copulas (normal, *t*, Gumbel, Frank) can be estimated - wind farm pairs that have other copulas selected have been omitted from the estimation process. Table 2 reflects the omission by only including wind farm pairs that have been estimated. Estimating the parameter means estimating the level of dependence between wind farms. Uncertainty in the association parameter means uncertainty in the dependence structure between wind farms. Such uncertainty will be reflected in the accuracy of the joint distribution that will be created from the copula and marginal distributions.

Aside from a posterior's confidence integral, the uncertainty can also be shown graphically when using the modified predictive posterior to forecast the wind speed for different wind farms. This predictive posterior is used to assess the accuracy of the joint distribution by comparing various simulated distributions to the original wind speed data for each wind farm. Once parameters are estimated, a looping method is used to create

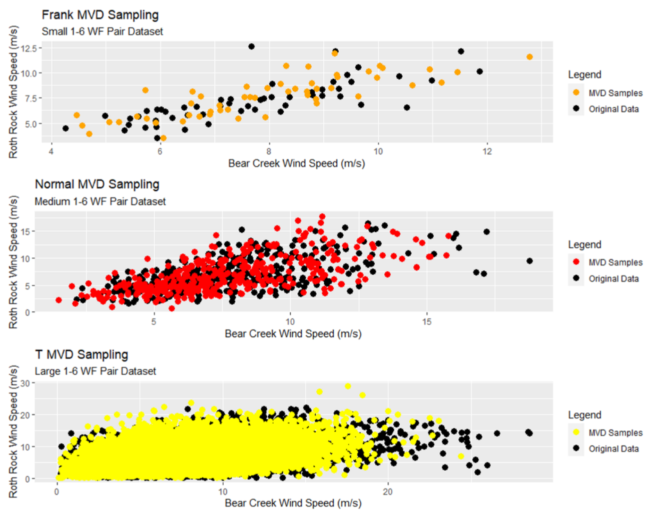

FIGURE 6 Point samples from the multivariate distributions created for the 1–6 wind farm pair. In these point plots, Bear Creek and Roth Rock are plotted against each other to form a two-dimensional plot of wind speed data. The black points represent original data while the coloured points represent simulated wind speed data from the multivariate distributions. The simulated data for weekly wind speed is very inaccurate compared to the simulated data for the larger sets

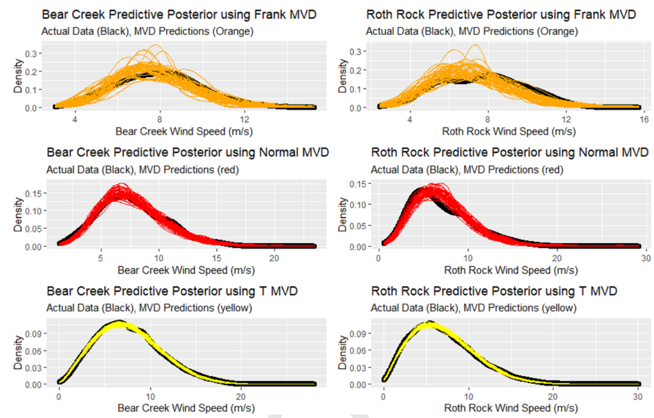

FIGURE 7 Predictive posterior for all datasets for the 1–6 WF pair data. Predictive posterior plots for the 1–6 wind farm pair under each data volume. These plots represent wind speed distributions sampled from the multivariate distributions (MVD). The top row show sampled weekly data from the Frank MVD, the middle row represents sampled data from the normal MVD and the bottom row represents data from the Student-t MVD. The MVD with the weekly data set cannot capture the original wind speed data as well as the other MVDs can

numerous distributions of predicted wind speed from temporary joint distributions using the estimated parameters from the posterior. These predictions are then compared to the original wind farm data for accuracy, and can later be used as input for power flow analysis.

4 | SIMULATION AND RESULTS

As the volume of wind speed data increases, the certainty in parameter estimation increases. For the medium data set, there is more certainty in parameter estimation than in the small data set. The copula certainty is the best using the hourly (8760) data set. The choices of copulas in Table 2 reflect strong dependence between wind farms due to the close spatial relationship. The normal, t, Frank, BB1, and BB8 copulas were selected to model the dependence structures of the wind farms. However, the copulas estimated under Bayesian Estimation are the normal, Student-t, and Frank copulas.

When examining the smaller data sets for each wind farm pair, there is a noticeable deviation from the posterior mean in every pair. The two copulas most used for wind farm pairs under the small data set were the normal and Frank copulas. There is not enough data to make out a structure to properly represent these copulas. In Table 2, the CI is wide, denoting much uncertainty about the ideal association parameter for each copula. For wind farm pairs 3–5, the normal copula represents the small volume of data. The CI is between 0.613 and 0.73 – a wide interval. For another pair, 1–6, the Frank copula is chosen to represent its small data. The CI is very wide, with values between 7.15 and 12.9 for the Frank copula parameter.

The uncertainty is also shown in the resulting multivariate distribution created from the copula for wind farm pair 1–6 in Figure 5. The multivariate distribution's forecasting ability is determined by the reiterative predictive posterior method, where each parameter value from the posterior is used to create multivariate distribution predictions to be compared to the original data. As shown in Figure 7, the predictive posterior shows that the distributions do not come close to predicting wind speed accurately using small data sets.

Under medium data sets, the CI for each posterior is much smaller compared to using a small data set. The copula selections are different as the normal copula was mostly chosen. Such a choice shows an equal amount of similarities between wind farms at the tails and on the diagonal. As the normal copula was chosen numerous times, the other copula chosen often was the BB1. Although the parameters of the BB1 copula could not be estimated using Bayesian inference, this copula shows that some wind farms have asymmetrical tail dependence between one another. The amount of data in the medium data set is enough to see a dependence structure between two wind farms. Due to more certainty in the copula parameter, the forecasting power of 1–6′s multivariate distribution in Figures 6 and 7 are much better. However, there is still a level of uncertainty that should be captured when using such a data set.

For the 8760 data, there is very little to no uncertainty in the copula parameter estimation. The CI is very small – resulting in the posterior mean equalling the initial chosen copula parameter from non-Bayesian methods. The choice of copulas is much different from the other data sets because of the large volume of data. Unlike the copulas found in the smaller data sets, the Student-t copula was the most selected. This copula captures extreme wind speed observations between two wind farms. Other copulas selected that model extreme tails to include the Tawn copula, BB1 and BB8 copulas. The large data set has more than enough data to model a noticeable dependence structure. The forecasting power of the multivariate distributions using this data set is nearly perfect, as shown in Figures 6 and 7. This shows that the larger data sets provide enough observations to safely use non-Bayesian techniques to estimate copula parameters.

5 | CONCLUSION

In this paper, Bayesian inference was used to estimate the copula parameter for each wind farm pair under weekly, daily, and hourly wind speed observations. The Bayesian approach was successfully utilized for elliptical and single-parameter Archimedean copulas. Due to the close proximity of each wind farm, strong tail and central dependence were captured using the normal, *t*, Frank, BB1 and BB8 copulas. In estimating the copula parameter for each wind farm pair, the posterior's CI was examined to determine the uncertainty in parameter estimation. The results show that a smaller volume of data will have more uncertainty in copula parameter estimation than 8760 hourly data points. From the results, we conclude that non-Bayesian parameter estimation can be used for hourly wind speed data with certainty in the estimated value.

FUNDING
None.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available in the well-known NREL at WWW.nrel.gov. These data were derived from the following resources available in the public domain: WWW.nrel.gov

ORCID
Amir Shahirinia https://orcid.org/0000-0002-0081-0674
M. Tavakoli Bina https://orcid.org/0000-0001-9344-146X

REFERENCES
1. Athari, M.H., Wang, Z.: Impacts of wind power uncertainty on grid vulnerability to cascading overload failures. IEEE Trans. Sustainable Energy 9(1), 128–137 (2017)
2. Wang, Y., Zhang, N., Kang, C., Miao, M., Shi, R., Xia, Q.: An efficient approach to power system uncertainty analysis with high-dimensional dependencies. IEEE Trans. Power Syst. 33(3), 2984–2994 (2017)
3. Shahryari, E., Shayeghi, H., Mohammadiivatloo, B., Moradzadeh, M.: A copula-based method to consider uncertainties for multi-objective energy management of microgrid in presence of demand response. Energy 175, 879–890 (2019)
4. Philippe, W.P.J., Eftekharnejad, S., Ghosh, P.K.: A Copula-based uncertainty modeling of wind power generation for probabilistic power flow study. In: 2019 IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE). Oshawa, ON, Canada, 12—14 August 2019
5. Wang, Z., Wang, W., Liu, C.,Wang, Z., Hou, Y.: Probabilistic forecast for multiple wind farms based on regular vine copulas. IEEE Trans. Power Syst. 33(1), 578–589 (2017)
6. Li, M., Lin, Z., Ji, T.,Wu, Q.: Risk constrained stochastic economic dispatch considering dependence of multiple wind farms using pair-copula. Appl. Energy 226, 967–978 (2018)
7. Ehsan, M., Shahirinia, A., Gill, J., Zhang, N.: Dependent wind speed models: Copula approach. In: 2020 IEEE Electric Power and Energy Conference (EPEC). Edmonton, AB, Canada, 9–10 November 2020
8. Stephen, B., Galloway, S.J., McMillan, D., Hill, D.C., Infield, D.G.: A copula model of wind turbine performance. IEEE Trans. Power Syst. 26(2), 965–966 (2010)
9. Cao, J., Yan, Z.: Probabilistic optimal power flow considering dependences of wind speed among wind farms by pair-copula method. Int. J. Electr. Power Energy Syst. 84, 296–307 (2017)
10. Philippe, W.P.J., Zhang, S., Eftekharnejad, S., Ghosh, P.K., Varshney, P.K.: Mixed copula-based uncertainty modeling of hourly wind farm production for power system operational planning studies. IEEE Access 8, 138569–138583 (2020)
11. Karakaş, A.M.: Using copulas for modeling dependence in wind power. Asian J. Eng. Technol. 7(1), (2019)
12. Shahirinia, A., Hajizadeh, A., David, C.Y.: Bayesian predictive models for Rayleigh wind speed. In: 2017 IEEE 17th International Conference on UbiquitousWireless Broadband (ICUWB). Salamanca, Spain, 12–15 September 2017
13. Monahan, A.H.: Idealized models of the joint probability distribution of wind speeds Nonlin. Processes Geophys. 25(2), 335–353 (2018)
14. Veeramachaneni, K., CuestaInfante, A., O'Reilly, U.M.: Copula graphical models for wind resource estimation. In: Proceedings of the 24th International Conference on Artificial Intelligence. Buenos Aires, Argentina, 25—31 July 2015 Q5
15. Ly, S., Pho, K.H., Ly, S., Wong, W.K.: Determining distribution for the product of random variables by using copulas. Risks 7(1), 23 (2019)
16. Sklar, M.: Fonctions de repartition an dimensions et leurs marges. Publ Inst Statist Univ, Paris 8, 229–231 (1959)
17. Schmidt, R.: Tail dependence. In: Statistical Tools for Finance and Insurance, pp. 65–91. Springer, Berlin Heidelberg (2005)
18. Shemyakin, A., Kniazev, A.: Introduction to Bayesian Estimation and Copula Models of Dependence. Wiley, Hoboken, New Jersey (2017)
19. Lin, Y., Yongning, Z.: A review on wind power prediction based on spatial correlation approach. Autom. Electr. Power Syst. 38(14), 126–135 (2014)
20. Silva Saravia, H., Pulgar Painemal, H.: Effect of wind farm spatial correlation on oscillation damping in the wecc system. In: 2019 North American Power Symposium (NAPS). Wichita, KS, USA, 13–15 October 2019
21. Joe, H., Hu, T.: Multivariate distributions from mixtures of max-infinitely divisible distributions. J. Multivariate Anal. 57(2), 240–265 (1996).

How to cite this article: Henderson S.B., Shahirinia A., Bina M.T. Bayesian estimation of copula parameters for wind speed models of dependence. IET Renew. Power Gener. 1–9 (2021) https://doi.org/10.1049/rpg2.12297.

APPENDIX
The BBx copulas are a set of bivariate Archimedean copulas that combine the properties of two single-parameter Archimedean copulas. The BB1 and BB8 will be reviewed as these are the two-parameter copulas chosen outside of the other copulas for this experiment.

BB1 copula

Referred to as the Clayton-Gumbel copula, the BB1 copula captures asymmetric tail dependence. The copula function for the BB1 is defined as:

$$C_{\theta\delta}(u, v) = (1 + ((u^{-\theta} - 1)^{\delta} + (v^{-\theta} - 1)^{\delta})^{\frac{1}{\delta}})^{\frac{-1}{\theta}} \tag{14}$$

where $\delta \geq 1$ and $\theta > 0$.

BB8 copula

The BB8 copula, or the Joe-Frank copula, captures upper tail and central dependence. The BB8 copula is defined as follows:

$$C_{\theta\delta}(u, v) = \frac{1}{\theta}\left(1 - \left(1 - \frac{1}{1 - (1 - \delta)^{\theta}}\left(1 - (1 - \delta u)^{\theta}(1 - \delta v)^{\theta}\right)\right)^{\frac{1}{\delta}}\right) \tag{15}$$

where $\theta \geq 1$ and $\delta \in [0, 1]$.