Energy Conversion and Management 252 (2022) 114950

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

A Bayesian model for wind farm capacity factors

David Wilkie [a,*], Carmine Galasso [a,b]

[a] Department of Civil, Environmental and Geomatic Engineering, University College London, London, United Kingdom
[b] Scuola Universitaria Superiore (IUSS) Pavia, Pavia, Italy

ARTICLE INFO

Keywords:
Offshore wind farm
Capacity factors
Levelized cost of energy
Bayesian model

ABSTRACT

Capacity factors are an important performance metric for offshore wind energy projects as they indicate how efficiently a given project generates electricity. Given the intermittent nature of the wind resource, there is substantial variability between observed capacity factors seasonally and between years. However, little work has focused on extracting trends in the variable wind farm energy generation data. This paper proposes applying hierarchical Bayesian techniques to historical capacity factors to enable the prediction of capacity factor distributions. The proposed model relies on data from UK offshore wind farms, the most developed market for offshore wind energy, but is equally applicable to other countries. The resulting capacity factor distributions highlight a substantial variability in capacity factors both when modelled as yearly and monthly (which accounts for seasonality). The model shows that newer wind farms have higher capacity factors than older farms, with an improvement of approximately 20%. It is also demonstrates that capacity factor variability has a smaller impact on levelized cost of energy estimates. The results from this study can be used both to predict capacity factors for individual wind farms or to make predictions for a generic wind farm.

1. Introduction

Offshore wind is a fast-growing form of electricity generation. Europe reached 22.5GW of installed capacity by the end of 2019 [1], increasing from only 0.9GW in 2010. This growth has been notable in the UK, which has approximately half of Europe's installed capacity. The scale of offshore wind farms (OWFs) has increased, as has the capacity of individual offshore wind turbines (OWTs) [1]. However, wind energy is intermittent as the electricity generated depends on the local wind climate. The electricity generation can be summarised through various metrics, including the annual energy productions (AEP), defined as the electricity generated over a year. Feasibility studies commonly use this metric for potential OWF sites (e.g., as is the case in Brower et al. [2] or Staffell et al. [3]) and also for the resale of operational wind farms [4]. However, the AEP is based on granular information about the site environmental conditions, such as interpolated wind speeds from National Aeronautics and Space Administration (NASA) climate reanalysis models and is therefore difficult to calculate. Capacity factors (also called load factors [5]) are another common performance metric. They represent the ratio of electricity generated to the maximum possible generation based on the installed capacity, i.e., calculated assuming that the asset functioned at its maximum capacity over the investigated period of time. Capacity factor estimates are essential for feasibility studies and the development of wind energy projects [6], such as in Atkin's net-zero by 2050 report [7].

Estimating energy production is essential not only on its own but also as an input to the levelized cost of energy (LCoE). This metric is used by the UK government [8] and agencies such as the International Renewable Energy Agency [9] to compare the cost of different electricity generation technologies. It is also used to assess OWF viability during the licensing process and operational phase to secure financing, refinancing, and potentially sale. Capacity factors are often used to evaluate the LCoE instead of the AEP, as demonstrated in a wide range of institutional reports, including that of the UK Government's Department for Business, Energy & Industrial Strategy (BEIS) [8], the Danish Energy Agency's review of generating costs [10], and the National Renewable Energy Laboratory's Annual technology baseline [11].

There has been a range of studies estimating various performance metrics for renewable energy projects through advanced statistical methods, including weather forecasting for renewable energy prediction. For instance, Altan et al. [12] developed a wind time-series forecasting model combining long short-term memory neural network, decomposition methods and a grey wolf optimiser. Karasu et al. [13] predicted solar radiation using a random forest and moving average

* Corresponding author.
E-mail addresses: David.wilkie.15@ucl.ac.uk (D. Wilkie), c.galasso@ucl.ac.uk (C. Galasso).

https://doi.org/10.1016/j.enconman.2021.114950
Received 12 May 2021; Received in revised form 26 October 2021; Accepted 27 October 2021
Available online 16 December 2021

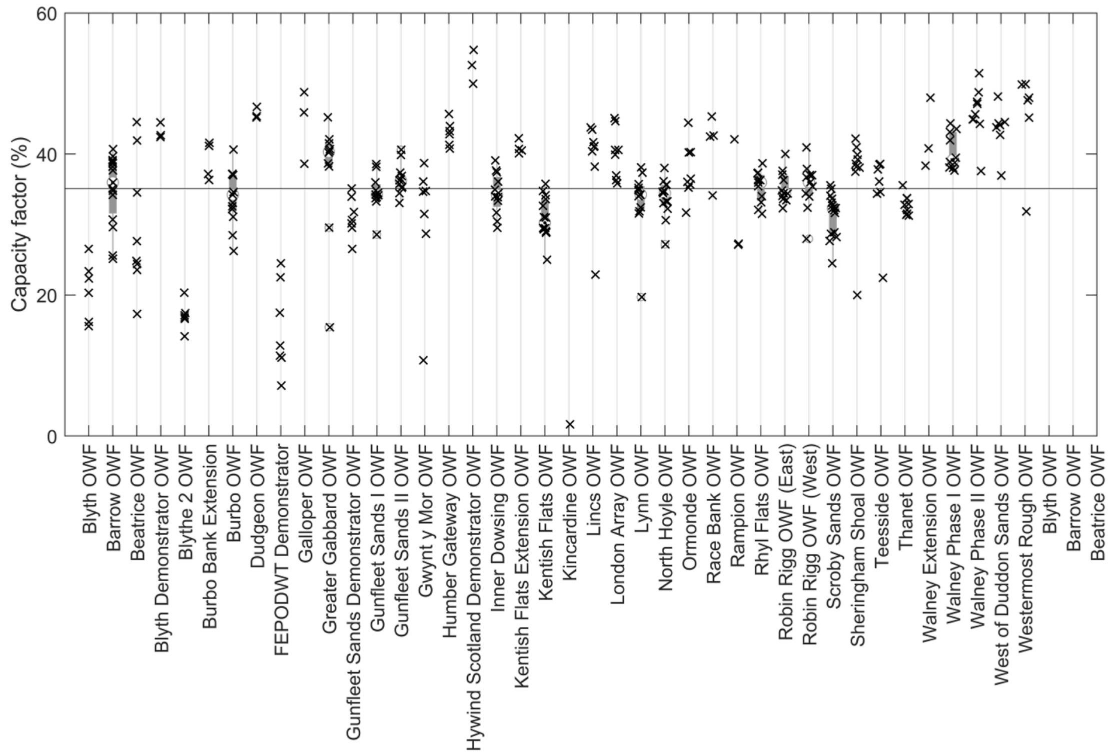

Fig. 1. Capacity factors calculated for UK offshore wind farms (OWFs). Box-plots are only shown for OWFs having over 10-years of operational data. The first year of production is taken as the first year in which there is more than half a year of generation data after the OWF is reported as fully commissioned in the 4C Offshore database [21]. The mean evaluated across all OWFs is indicated with a solid black line.

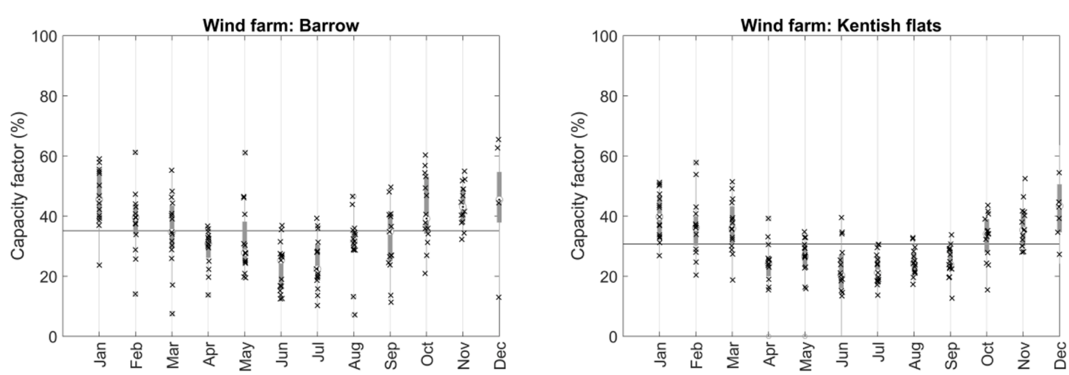

Fig. 2. Capacity factors calculated monthly for the Barrow (left) and Kentish Flats (right) OWFs. The mean evaluated across the months of operation is indicated with a solid black line.

model. Models have also been developed to wind energy prediction by Bodini [4] for US onshore turbines using a correlation-based model. In the literature, a range of approaches has been used to estimate capacity factors. Some studies use qualitative predictions of the capacity factor based on assumptions about OWFs. For instance, the BEIS assumes a value of 43% for Round 2 OWFs (i.e., farms announced in 2017) and 48% for Round 3 (i.e., farms announced in 2018) [8]. In contrast, Atkins [7] assume a value of 58% as a target for future developments, a 31% capacity factor for old OWFs or a 44% value for newer OWFs. Another approach involves averaging measured capacity factors; for instance, the yearly Digest of UK Energy Statistics (DUKES) for 2019 estimates a value of 40.1% by aggregating all UK OWFs [14]. Similarly, Ayodele et al. [15] calculated seasonal capacity factors for wind turbines in South Africa using measured data. Crabtree et al. [16] calculated capacity factors for all UK OWFs in 2015 by averaging measured energy production data, finding that Round 1 (i.e., announced in 2015) and Round 2 farms had an average capacity factor of 33.6% and 38.3%, respectively. Aldersey-Williams et al. [17] used the mean observed capacity factor over the operating life of the OWFs in their study. However, for OWFs with short production histories, they assume a value of 48%. The only source of capacity factors that is up to date and covers individual UK OWFs comes from Energy Numbers [18].

Crucially then, the existing literature on capacity factors relies on either qualitative/subjective judgements or point estimates. Additionally, the existing work does not explicitly consider variability of the estimates, which would indicate the scatter in capacity factor. Indeed, relatively high variability in yearly capacity factors is noticeable when an OWF's operating history is plotted as a scatter chart, as shown in Fig. 1 (the data used to generate this plot is described in Section 2.1) for the UK's 43 operating OWFs. For example, both the Barrow and Kentish Flats OWFs have a large coefficient of variation of 11% in terms of capacity factor. In addition, there is also considerable variability in capacity factors seasonally, as demonstrated in Fig. 2. In this context, Bayesian models provide an appropriate way to assess capacity factors.

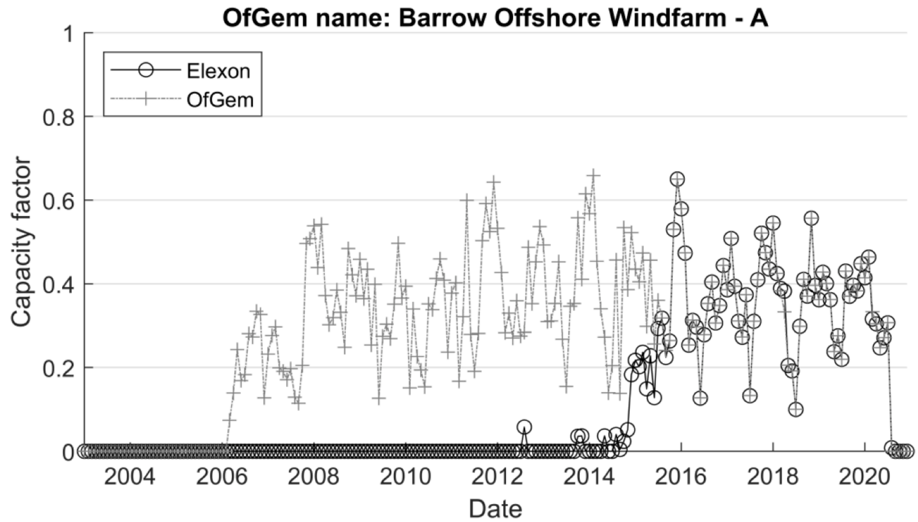

Fig. 3. Comparison between the monthly capacity factor calculated for the Barrow OWF using the OfGem data (dotted line with a plus sign) and the Elexon data (solid line with a circle).

They have seen many applications, including predicting unknown quantities in engineering problems, such as the properties of structural systems [19] and cost data for renewable projects [20]. This paper proposes using a hierarchical Bayesian model to estimate capacity factors for UK OWFs, providing a novel flexible model that can predict the capacity factors for both an individual OWF and groups. This model can be used to estimate mean capacity factors and also evaluate variability in their estimation. The model works for both annual and monthly capacity factors. This study focuses specifically on UK data to illustrate/test the proposed model because the country is the most mature market for OWFs. However, the model developed in this study could be applied to other countries for which historical capacity factor data is available.

Section 2 presents the data related to OWF electricity generation and then the cost data required to evaluate the LCoE. It also contains a summary of the Bayesian algorithms utilised in this study. Section 3 includes a representative sample of the results and a discussion of the model quality of fit. The full results for each of the 24 UK OWF in this study are presented in the Appendices/Supplementary data to this article.

2. Data and methods

This section starts by summarising the data used in the study. Then the proposed Bayesian model is introduced, along with the tests used to verify its accuracy. Finally, the calculation used to evaluate the LCoE is presented, enabling determining the impact that uncertainty in capacity factors have on cost.

2.1. Capacity factor data

The capacity factor (CF) for an OWF is calculated by dividing the electricity generated E (in MWh) over a given period h (in hours) by the electricity that could have been generated based on the installed capacity [14]:

$$CF = \frac{E}{0.5(C_o + C_e)h} \tag{1}$$

where C_o is the installed capacity at the start of the period (in MW) and C_e is the installed capacity at the end of the period (in MW). The capacity factor is calculated in this study using two different sources of information and compared to ensure that the values predicted are correct.

The first source of electricity generation information is the Office of Gas and Electricity Markets (OfGem) [22] which holds a database of large-scale renewable energy assets commissioned before 2019 and which also contains details of the installed capacity. For this study, all records from January 2000 to December 2020 were used. OfGem issues Renewable Energy Guarantees of Origin (REGO) for each unit of electricity generated by renewable energy sources in their database, allowing the actual generation of the UK's OWFs to be evaluated. The certificates are issued in batches, and each refers to a single unit of electricity generation; specifically, before 2010, each batch represented one kWh, whereas those issued after 2010 represented one MWh. The certificates were processed so that those batches of certificates covering multiple months were split evenly between the relevant months. For those months in which multiple batches of certificates were issued, the total electricity was calculated by summing the relevant batches. The REGO allows renewable energy generation to be traded between companies; therefore, the certificates are assigned a status depending on how the electricity has been processed. This calculation includes all certificates that have one of the following states:

- Issued - after OfGem confirms generation;
- Redeemed - a status given once an issued REGO is utilised within a suppliers fuel mix disclosure;
- Retired - used for Northern Ireland where OfGem does not redeem the certificates;
- Expired - when a REGO related to electricity generation more than 16 months in the past.

Additionally, certificates can be assigned a Revoked status, meaning that OfGem has determined the REGO request to be false or inaccurate [23]. These were not used to calculate generation and accounted for only 0.5% of entries in the OfGem database. The full database processed as described above is made available by the authors through GitHub [24].

The second source of electricity generation information is Elexon, a private company that manages electricity trading within the UK. The company's remit includes metering the production at generating facilities. They maintain a metered half-hourly output database from a range of UK large-scale electricity generating assets, including many OWFs [25]. Production data for each OWF is accessible through an Application Programming Interface (API) and was converted into monthly production values by summing each half-hour across the relevant month. The data was cleaned by removing data that was not a number, but no other processing was applied.

The monthly energy production values and resulting capacity factors

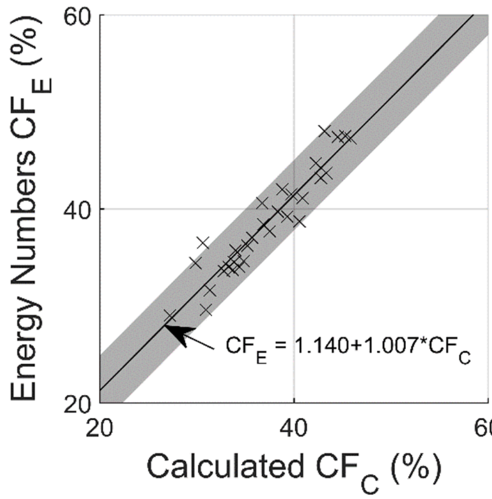

Fig. 4. Comparison between the lifetime average capacity factor calculated by the Energy Numbers website (y-axis) and the analysis in this paper (x-axis). The black line indicates a linear relationship fit to the data, and the grey area represents the 95% confidence interval.

calculated using the OfGem and Elexon database were compared to verify that both sources produced similar results. A typical comparison is shown in Fig. 3 for the Barrow OWF. This figure shows that the Elexon database does not have metered data properly registered before 2016, although it is known from the 4C database that the Barrow OWF was fully commissioned in 2006 [21]. However, recent years show a very good agreement between the data sources. This validates the post-processing applied to the OfGem database, namely extracting monthly production data from certificates that can cover multiple months, and confirms that the Elexon database is accurate for recent data.

Additionally, the capacity factor calculation was verified by comparing the individual lifetime capacity factor calculated in this study to those on the Energy Numbers website [18] for 44 operational UK OWFs. The Energy Numbers website did not provide a full history of calculated capacity factors but only an annual summary; therefore, it could not be used as a data source. The results are shown in Fig. 4 and indicate that the two calculation approaches produce similar capacity factors. The yearly capacity factors were calculated by taking the mean over the monthly capacity factors and assuming the year began in January and ended in December. The first year of operation was taken to be the first full year since commissioning or the first year with more than six months of operating data (in this case the expectation was calculated July-December to prevent bias caused by the reduction in capacity factor towards the summer months observed in Fig. 2).

2.2. Cost data

OWF costs are split between capital and operational expenditure. The capital expenditure is the fixed cost of the physical assets, which includes the turbine, the foundation, and the installation. The operational expenditure is the cost of generating energy; it is split between fixed expenditure, occurring independently of the amount of electricity generated, and variable expenditure that depends on the electricity generated (such as maintenance or fuel costs).

The capital cost data is taken from Aldersley-Williams et al. [26], who evaluated capital costs using published accounts and are summarised in Table 1 for the 23 OWFs in their study. The capital expenditure is distributed over the construction of the OWF. A three-year construction period is assumed where the capital expenditure is 60% for the first year, 30% for the second year and 10% for the third year [11].

A constant value of the fixed operational expenditure of 70,000 (in pound sterling/MW or £/MW) and operational expenditure of 4.7 (£/MWh) are assumed for this study. These values were selected to make

Table 1
Total capital expenditure costs used in this study, where the capital expenditure data is taken from Aldersey-Williams [26]. Note: for the farm with an asterisk (*), the relevant values are used for the yearly model but not the monthly model due to the small quantity of observed data for some months.

Farm	Capital expenditure (£ million/MW)	Date of first production	Installed Capacity (MW)	LCoE (£/MWh)
Barrow	1.49	2006	90	87.15
Burbo Bank Extension*	4.05	2016	90	146.95
Burbo Bank	1.79	2007	259	86.89
Dudgeon*	2.09	2017	402	104.07
Greater Gabbard	3.21	2011	504	136.62
Gwynt y Mor	3.41	2014	576	179.18
Humber Gateway	3.48	2015	219	147.13
Inner Dowsing	2.15	2008	97.2	96.59
Kentish Flats	1.22	2006	139.5	66.43
Lincs	3.66	2013	270	166.053
London Array	3.13	2012	630	139.89
Lynn	2.22	2008	97	101.54
North Hoyle	1.88	2003	60	77.35
Ormonde	3	2011	150	149.08
Rhyl Flats	2.84	2009	90	125.62
Robin Rigg	2.79	2010	174	135.49
Scroby Sands	2.3	2006	60	104.83
Sheringham Shoal	3.78	2012	316.8	150.34
Teesside	3.53	2013	62.1	235.96
Thanet*	3.14	2010	300	158.69
Walney	2.89	2011	1026.2	120.37
West of Duddon Sands	1.5	2014	388.8	72.11
Westermost Rough*	2.58	2014	210	120.82

the LCoE predictions from this study match those in Aldersey-Williams et al. [26] (evaluating the LCoE using the mean lifetime capacity factor). This is a simplification as there will be a unique value for each OWF depending on its distance to shore, among other factors. However, the operational expenditure is small compared to the capital expenditure [11]; in addition, this study aims to determine the variability in LCoE caused by modelling the capacity factor probabilistically and not comparing the LCoE between UK OWFs.

The LCoE is sensitive to the discount rate [27], where the weighted annual cost of capital is commonly used in the literature as it provides a balance between capital financed between equity and debt. The BEIS [8] use a real rate of 8.9% for offshore wind projects in the UK, which has also been adopted in this study, representing the weighting between the portion of capital expenditure financed by external debt (borrowing) and by internal debt (equity). Furthermore, an operational life of 25 years is assumed for every OWF.

2.3. Proposed Bayesian capacity factor model

Hierarchical Bayesian models provide a general framework enabling a parameter of interest's variability to be quantified and parameterised. This technique is suitable to model the variability in capacity factors because it explicitly captures the parameters of interest as probability distributions to account for parametric uncertainty. A complete introduction to Bayesian methods can be found in Gelman et al. [28]. These models predict a posterior distribution for the parameter of interest by utilising an assumed prior (in the form of a prior distribution) and conditioning it on observed data (through a likelihood function). In addition, a hierarchical model adds tiers to the prior, linked through a statistical relationship, thereby introducing additional flexibility. The complete model can be thought of as a structured mathematical model, or graph, comprised of nodes connected through a probabilistic relationship that is one-directional and has no loops. This inherent flexibility makes them a suitable tool for modelling OWF capacity factors. In fact, there is an underlying similarity between OWFs from the same round, as they rely on similarly sized turbines and differences due to their location.

Figs. 1 and 2 highlighted variability in capacity factors calculated both in a yearly and monthly fashion. Therefore, two models are developed for this study, one incorporating yearly capacity factors and one considering capacity factors evaluated monthly, allowing a more detailed evaluation. The capacity factor component of this study uses 24 UK OWFs, which excludes only demonstrator projects and those farms with less than one-years worth of operating data in the OfGem database (the full list is provided in Table 2).

The yearly capacity factor model (CF_{kji}) in % predicts a capacity factor indexed by i for each of J offshore wind farms that are assigned to one of K licensing rounds, shown schematically in Fig. 5 (left). The likelihood function was selected to be a Normal probability density function (p) with parameters $\theta_{\mu,kj}$ for the mean and $\theta_{\nu,kj}$ for the standard deviation. The level 2 parameters relate to the wind farm licensing round and provide the mean of the wind farm distribution $\phi_{\mu,k}$ and the standard deviation of the wind farm distribution $\phi_{\nu,k}$. The level 3 hyper-priors have a Normal distribution with mean and standard deviation ($\boldsymbol{\tau}$). A non-informative distribution is used to represent the hyper-priors with a uniform distribution over the limits 0 to 100 (as the capacity factor is limited between 0% and 100%). The resulting posterior distribution defined by:

$$p(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\tau}|\boldsymbol{CF})\propto \mathrm{p}(\tau,\mu)\prod_{k=1}^{K}p(\phi_{\mu,k}|\tau_1,\tau_3^2)p(\phi_{\nu,k}|\tau_2,\tau_5^2)\prod_{k=1}^{K}\prod_{j=1}^{J}p(\theta_{\mu,kj}|\phi_{\mu,\mathrm{k}},\tau_4^2)p(\theta_{\nu,kj}|\phi_{\nu,\mathrm{k}},\tau_6^2)\boldsymbol{I_k}(j)$$

$$\prod_{k=1}^{K}\prod_{j=1}^{J}\prod_{i=1}^{n_i}p(CF_{kji}|\theta_{\mu,kj},\theta_{\nu,\mathrm{kj}})\boldsymbol{I_k}(j) \tag{2}$$

In Eq. (2), each OWF has $\boldsymbol{n_i}$ yearly capacity factors recorded and an indicator function $\boldsymbol{I_k}(j)$ is used to designate which j^{th} wind farm belongs to the k^{th} licensing round. The posterior distribution has $2J+2K+6$ dimensions. There were a total of 231 recorded yearly capacity factors for the 24 OWFs evaluated in this subset. This means that each sample of the from this posterior distribution requires $NK+2N+231$ probability estimates.

The monthly capacity factor model is shown in Fig. 5 (right), where the indices are the same as in Eq. (2) except for the addition of L, which refers to the number of months. The structure of this model was simplified in comparison to the yearly case by holding the variance fixed over all OWFs to reduce the number of parameters. The capacity factor distribution has a mean $\theta_{\mu,\mathrm{klj}}$ for each farm. This variable is distributed with a mean $\phi_{\mu,\mathrm{kl}}$ depending on the month and the standard deviation depending on a hyper-prior. The monthly mean also has a prior depending on the licencing round $\alpha_{\mu,\mathrm{k}}$, which is distributed according to hyper-priors. The posterior distribution is defined by the equation:

$$p(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\tau}|\boldsymbol{CF})\propto \mathrm{p}(\tau,\mu)\prod_{k=1}^{K}p(\alpha_{\mu,k}|\tau_1,\tau_2^2)\prod_{l=1}^{L}p(\phi_{\nu,l}|\tau_2,\tau_6^2)\prod_{k=1}^{K}\prod_{l=1}^{L}p(\phi_{\mu,kl}|\alpha_k,\tau_4^2)$$

Table 2
Summary statistics for yearly capacity factors predicted using the yearly and monthly models.

		Yearly		Monthly	
Farm	Round	Mean	Standard deviation	Mean	Standard deviation
Barrow	1	35.69	4.86	34.28	3.18
Burbo Bank	1	34.34	3.95	33.42	3.47
Inner Dowsing	1	35.59	4.65	32.82	3.18
Kentish Flats	1	31.87	3.90	34.87	3.42
Lynn	1	35.20	4.47	34.71	3.15
North Hoyle	1	33.86	3.09	33.55	3.32
Ormonde	1	37.90	5.32	34.68	3.58
Rhyl Flats	1	35.00	4.08	32.78	3.16
Robin Rigg (East)	1	35.44	3.52	33.74	3.23
Robin Rigg (West)	1	36.23	3.54	33.53	4.02
Scroby Sands	1	32.02	4.04	35.09	3.17
Teesside	1	35.05	7.01	34.28	3.52
Greater Gabbard	2	41.59	3.73	34.15	3.44
Gwynt y Mor	2	37.38	5.46	34.10	3.94
Humber Gateway	2	42.11	5.30	34.29	3.36
Lincs	2	41.09	2.69	34.90	3.34
London Array	2	40.80	5.51	39.84	5.09
Sheringham Shoal	2	39.61	2.11	33.52	3.30
Walney Phase I	2	40.22	2.75	32.99	3.44
Walney Phase II	2	45.25	4.68	34.46	3.10
West of Duddon Sands	2	44.27	2.65	40.05	4.93
Average		37.77	4.14	34.58	3.51
Round 1 Average:		34.88	4.25	33.95	3.35
Round 2 Average:		41.20	4.00	35.33	3.69

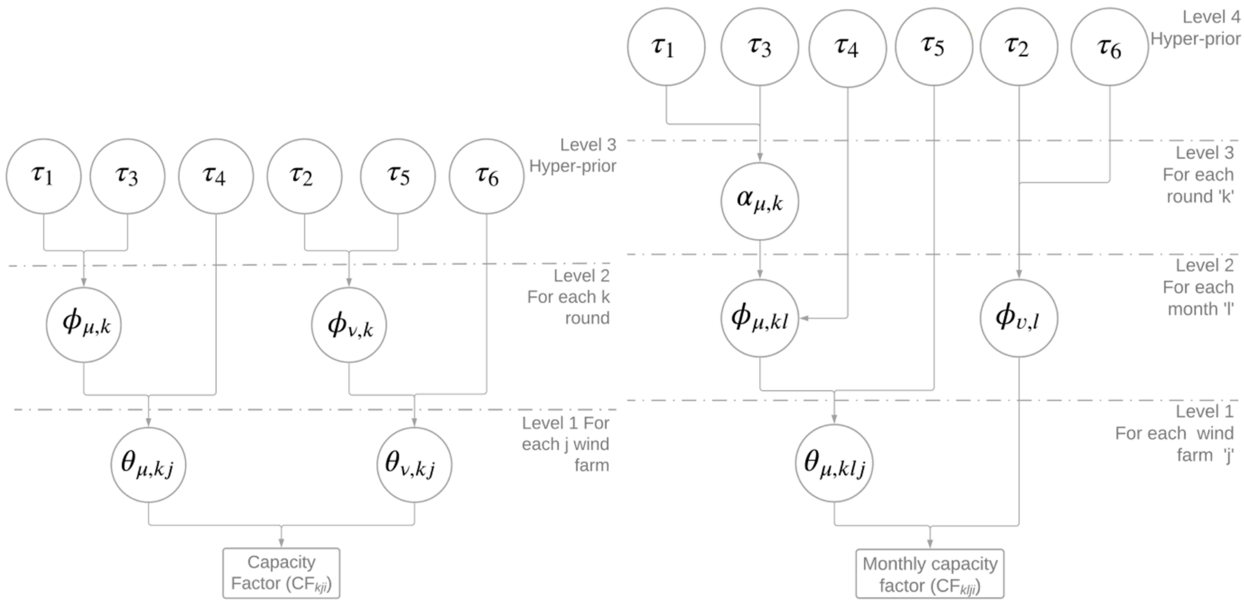

Fig. 5. Structure of the yearly (left) and monthly (right) hierarchical model for capacity factors.

$$\prod_{k=1}^{K}\prod_{l=1}^{L}\prod_{j=1}^{J}p(\theta_{\mu,klj}|\phi_{\mu,kl},\tau_5^2)\boldsymbol{I}_k(j)\prod_{k=1}^{K}\prod_{l=1}^{L}\prod_{j=1}^{J}\prod_{i=1}^{n_i}p(CF_{klji}|\theta_{\mu,klj},\phi_{\nu,l})\boldsymbol{I}_k(j) \tag{3}$$

In this model, all the distributions of the hyper-prior ($p(\tau,\mu)$) are modelled as a uniform distribution over the limits 0 to 100, and all other distributions are Normal. The posterior distribution has $JL+(1+K)L+K+6$ dimensions. There were a total of 2707 recorded monthly capacity factors for the 24 OWFs evaluated in this subset. This means that each sample from the posterior distribtuion requires $JL+(1+K)L+K+6+2707$ probability estimates.

In both cases, the posterior was simulated using Markov Chain Monte Carlo (MCMC) sampling implemented using the Metropolis-Hastings algorithm [28], with 12 chains and 17,500 samples each. The first 1,500 samples from each chain were removed to account for burn-in.

The calculations were implemented using a custom Matlab function using an adaptive method for the MCMC described by Haario et al. [29]. It calculates the empirical covariance matrix from the proceeding accepted samples and modifies this by a multiplicative tuning parameter. Additionally, a custom efficient algorithm was used to sample the Normal probability density function that does not rely on the error function and instead uses an iterative calculation based on Taylor expansion [30]. The calculation functions and input data are available via GitHub [24].

It should be noted that other combinations of prior models were tested, including Chi-squared, Student's and Lognormal distributions. This will not be discussed in the remainder of the paper, as the Normal distribution model was found to work best. Additionally, some theoretical justification for using a Normal distribution arises from the fact that energy generation is an additive process in the sense that the monthly generation is the accumulation (integral) of power.

2.3.1. Convergence

The convergence of the MCMC chains was assessed using the Multivariate Potential Scale Reduction Factor (MPSRF) [31], which is defined as the ratio between the total variance of the entire sample and the within-sequence variance. The within-sequence variance is evaluated by splitting the combined sample, run on multiple cores, into a number of segments. It is evaluated as:

$$\widehat{R} = \max_{a}\frac{\boldsymbol{a}^T\widehat{\boldsymbol{V}}\boldsymbol{a}}{\boldsymbol{a}^T\boldsymbol{W}\boldsymbol{a}} \tag{4}$$

Where $\widehat{V}$ is the total variance, $\boldsymbol{W}$ is the within sequence variance, and $\boldsymbol{a}$ is the vector that maximises $\widehat{R}$. When the MCMC has covered the sample space well, the MPSRF converges to 1, indicating that there is little between-sequence variance.

2.3.2. Model checking

Model checking is based on the Gelman et al. [28] Bayesian *p*-value (p_B), based on a generic test statistic ($T(y,\theta)$) which measures some aspect of the data that is valuable for the analysis, such as, for example, an inter-quantile range of the data. In a Bayesian context, the test statistic can be any useful transformation of the data. The Bayesian *p*-value is defined as the probability that the replicated data could be more extreme than the observed data, as measured by the test quantity. The *p*-value is defined:

$$p_B = \Pr(T(y^{rep},\theta) \geq T(y,\theta)|y) \tag{5}$$

which is based on comparing the replicated data (y^{rep}) to the observed data (y), potentially also depending on the posterior parameters (θ). This can be evaluated by integrating over the posterior distribution:

$$p_B = \int\int I_{T(y^{rep},\theta)\geq T(y,\theta)}p(y^{rep}|\theta)p(\theta|y)dy^{rep}d\theta \tag{6}$$

In practice, the integrals are solved numerically by drawing samples of replication data from the posterior distribution of parameters and evaluating the proportion of samples at which the test quantity for the replication data exceeds that form the observed data (i.e., for which $T(y^{rep},\theta) \geq T(y,\theta)$).

In this study, the test quantities used are:

- Minimum value;
- Maximum value;
- Mean value;
- Standard deviation.

2.4. Levelized cost of energy calculation

The *LCOE* in £/MWh is a key performance metric that depends on both electricity generation and cost data. It is defined as the "*cost that, if assigned to every unit of energy produced (AEP_n), will equal the total life-*

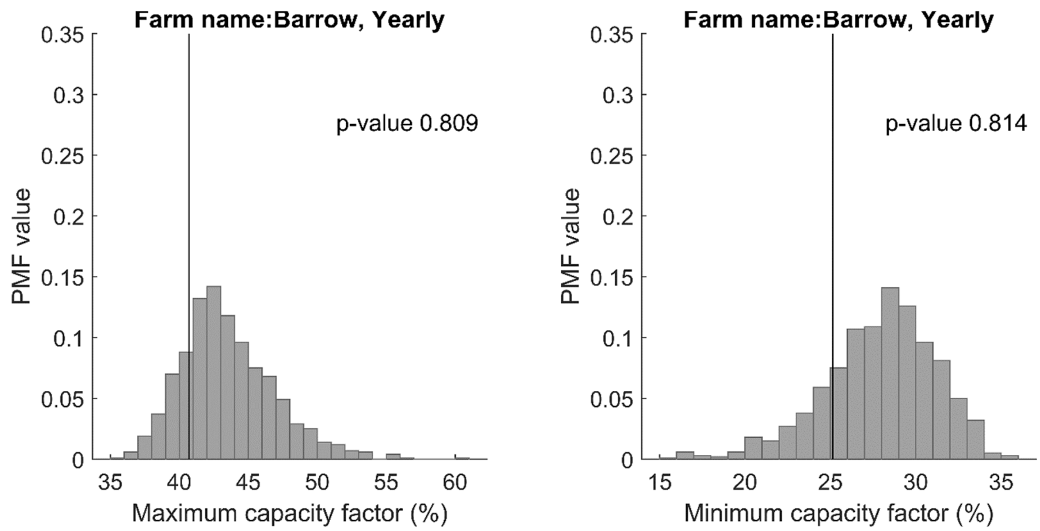

Fig. 6. Model checking results for the Barrow OWF.

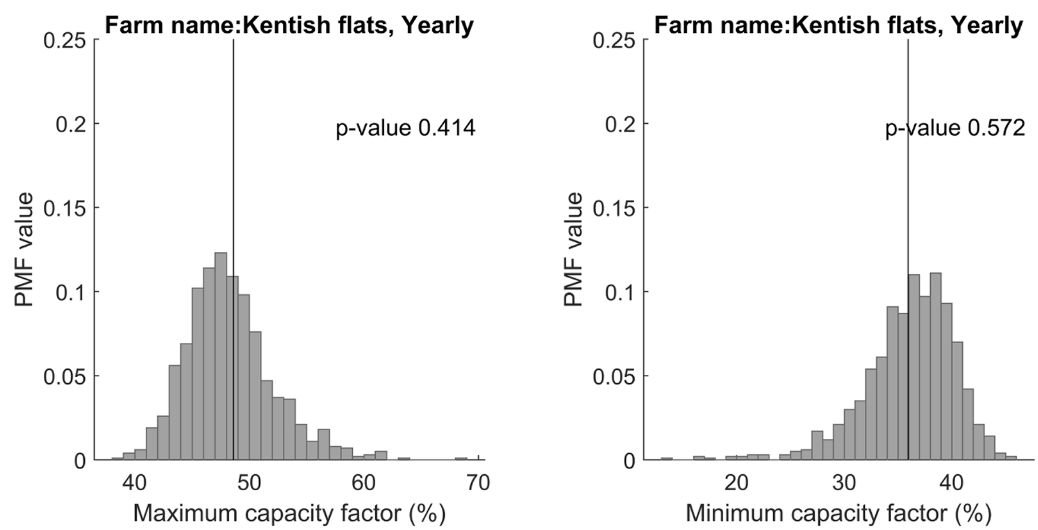

Fig. 7. Model checking results for the Kentish Flats OWF.

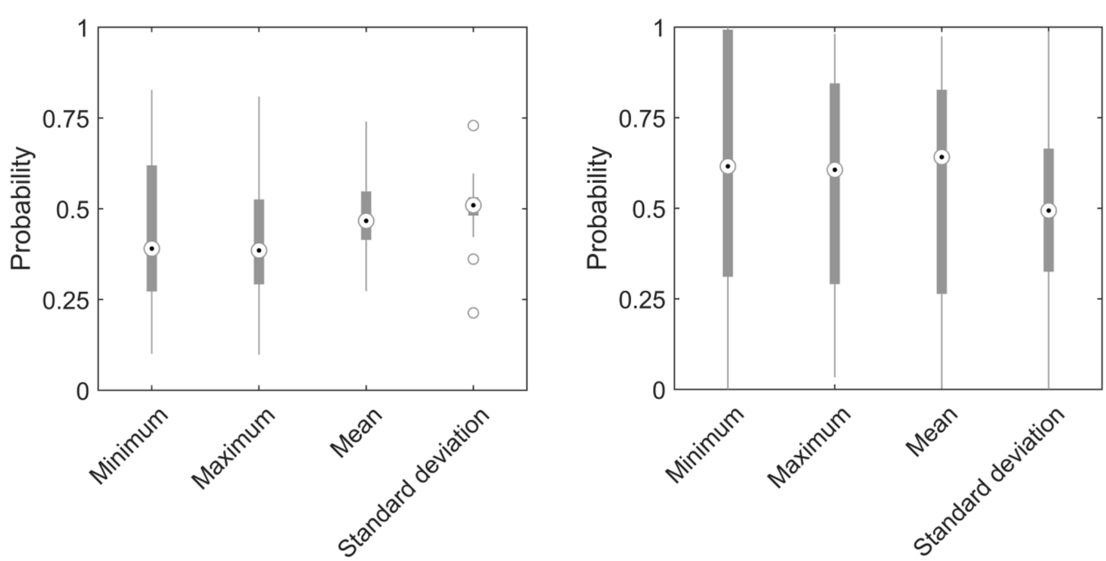

Fig. 8. Model checking results for the yearly model (left) and monthly model (right). The dots on the standard deviation plot are outlying values.

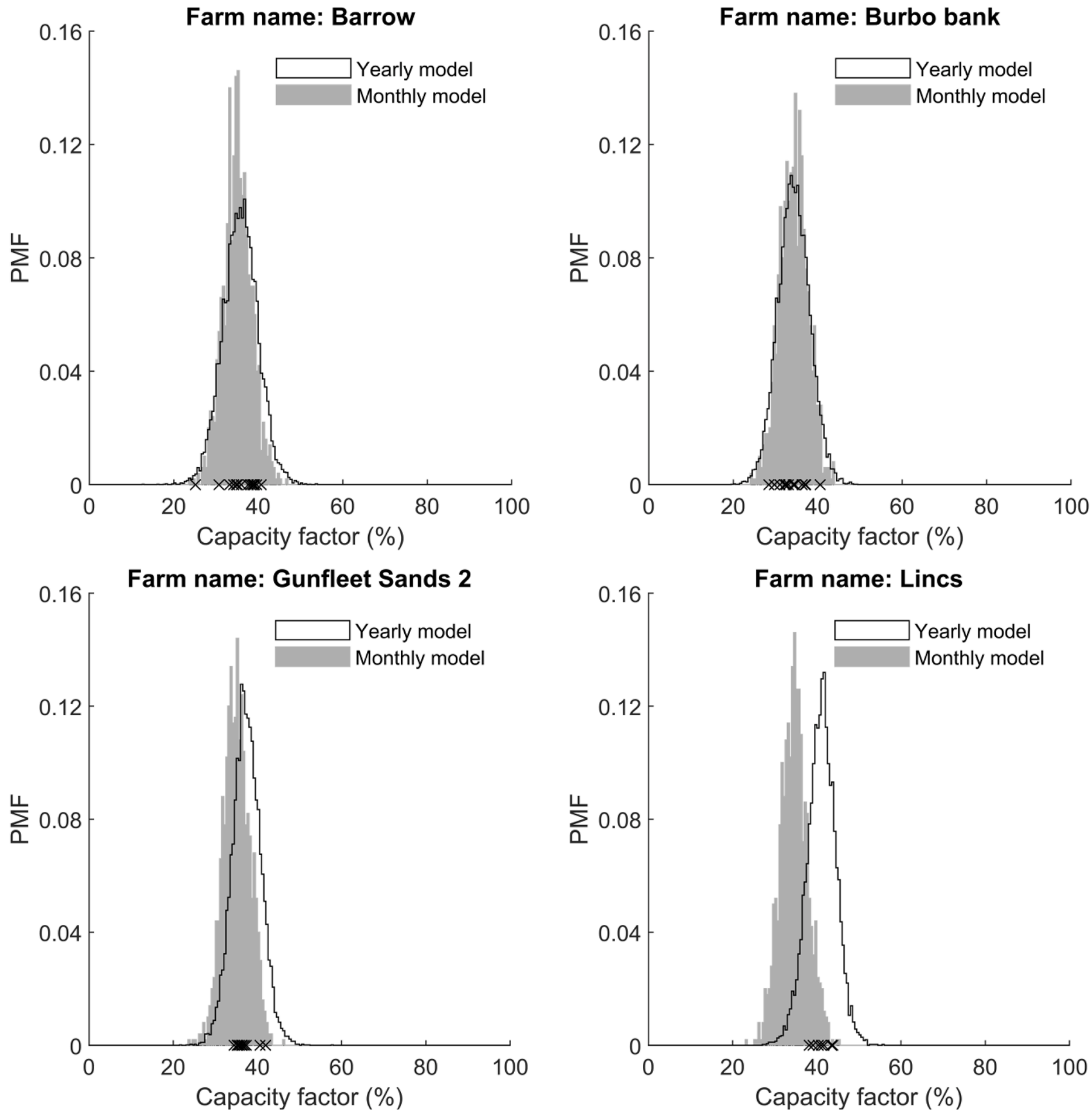

Fig. 9. Samples drawn from the posterior of the Bayesian model for the different Round 1 farms (top) and Round 2 farms (bottom). 'x' markers indicate the observed data for the OWF.

cycle cost (TLCC) when discounted back to the base year", where the *TLCC* is defined in £. It was originally defined by Short et al. [32]:

$$\sum_{n=1}^{N}\frac{AEP_{n}LCOE}{(1+r)^{n}} = TLCC \tag{7}$$

Eq. (7) can be rearranged to solve for *LCOE*:

$$LCOE = \frac{TLCC}{\sum_{n=1}^{N}\frac{AEP_{n}}{(1+r)^{n}}} = \frac{\sum_{n=1}^{N}\frac{\Delta C_{n}+O_{n}+V_{n}}{(1+r)^{n}}}{\sum_{n=1}^{N}\frac{AEP_{n}}{(1+r)^{n}}} \tag{8}$$

where the total life cycle costs are comprised of capital costs (ΔC_n) in £/MW, fixed operating costs (O_n) in £/MW, and variable operating costs (V_n) are discounted to the year of the analysis using the rate (r). These values may vary over the N years in the pay-back period, indexed by n.

The AEP_n depends on the capacity factor through the relationship [11]:

$$AEP_{n} = 8766CF_{n}C(1-AEP_{Loss}) \tag{9}$$

where CF_n is the capacity factor modelled using a Hierarchical Bayesian model (as described in the following section), AEP_{Loss} are the losses in energy production and C is the installed capacity of the OWF.

3. Results and discussion

The proposed Bayesian model is evaluated using the tests proposed in Section 2.3.1 and 2.3.2. Then the yearly and monthly model predictions are compared. Finally, the LCoE is calculated for the OWFs for which cost data were available using capacity factor predictions from the Bayesian model. These are compared and computed for a generic Round 1 or 2 wind farm in Section 3.3.

3.1. Model checking

The yearly model converged well with a maximum $\widehat{R}$ of 1.22 and a mean of 1.04. The monthly model has a maximum $\widehat{R}$ of 1.55 (for the τ_3 dimension) and a mean of 1.15.

The Bayesian p-value for the yearly model of the Barrow OWF is plotted in Fig. 6 as a representative example of the others. In addition, Fig. 7 shows the same representation for Kentish Flats, which has a p-value closer to 0.5 for the maximum and minimum capacity factor test. In these plots, the histograms are normalised into probability mass functions (PMFs). They show that resamples from the posterior are distributed well about the values in the observed data. In the yearly

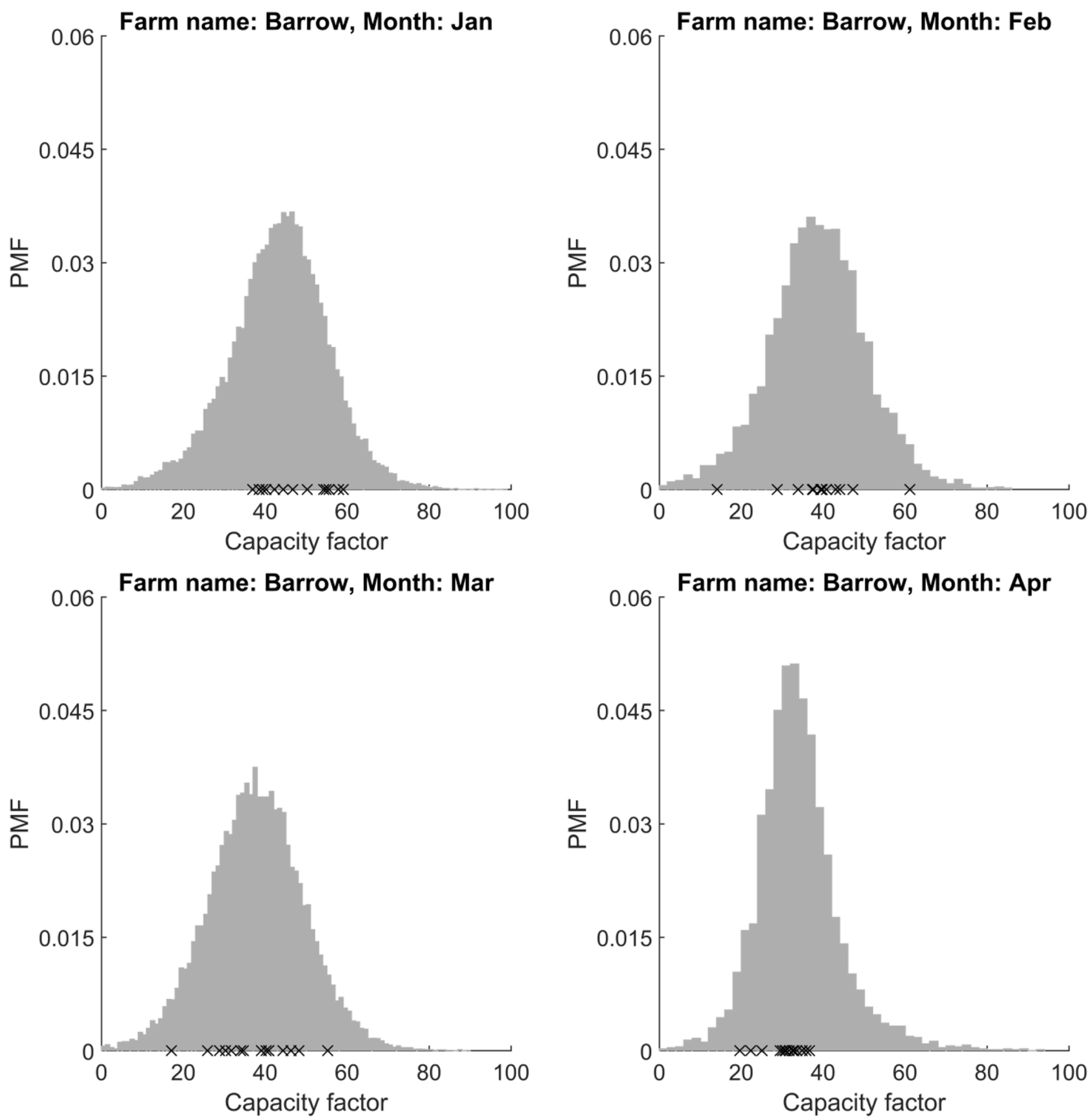

Fig. 10. Samples drawn from the posterior of the Bayesian model for different months at the Barrow OWF. 'x' markers indicate the observed data for the OWF in the month specified.

model, no *p*-values have an extreme value, taken as above 0.95 or below 0.05. This is further indicated in Fig. 8 (left), which shows the *p*-value as a box-plot over all the OWF in the model. A similar box-plot summarises the monthly model in Fig. 8 (right) over both OWFs and months; in this case, a total of 131 out of 1152 test values were found to hold extreme values. The minimum test caused 61 of these due to the model consistently predicting the minimum in the sample to be larger than that in the observed data. This is also suggested in Fig. 8 (right), where the mean of the minimum test box-plot is lower than 0.5.

Additionally, the mean test had 32 extreme values and tended to under-predict the observed sample mean. The extreme standard deviations were a mixture of under and over predictions. A full table of the results is provided in Appendix B for both the monthly and yearly models.

3.2. Model predictions

The posterior samples were used to generate predictive observations for both the monthly and yearly models. To compare both models, the monthly model capacity factors were converted into yearly capacity factors by taking the average over a years' worth of predictions. The mean of the posterior distribution samples for each farm in both the yearly and monthly models are reported in Table 2. Summary histograms of the yearly capacity factors predicted by both models are plotted against the observed yearly samples in Fig. 9 as examples of representative farms from Round 1 and Round 2 (the full set of plots for each OWF can be found in Appendix C). In these plots, the distribution of samples matches the observed data well, although for Barrow (Fig. 9 (left)) the monthly model appears to underestimate the distribution of samples. A histogram of predictions drawn directly from the monthly model is shown in Fig. 10 against the monthly capacity factor observations, showing a larger standard deviation in the monthly capacity factors than observed. This contrasts with the observations in Fig. 9 where monthly models had similar or reduced standard deviation compared to the yearly model but can be explained by the variance reducing effect of averaging over the year to produce Fig. 9.

For Round 1 OWFs, the monthly and yearly models predict similar yearly capacity factors. However, for the Round 2 OWFs, the monthly model predicts lower capacity factors as indicated by the averages at the bottom of Table 2 (41.20% vs 35.33%). This is primarily a consequence of the fewer observations in the monthly model, meaning that the more recent set of observations for the Round 2 farms have less "weight" than

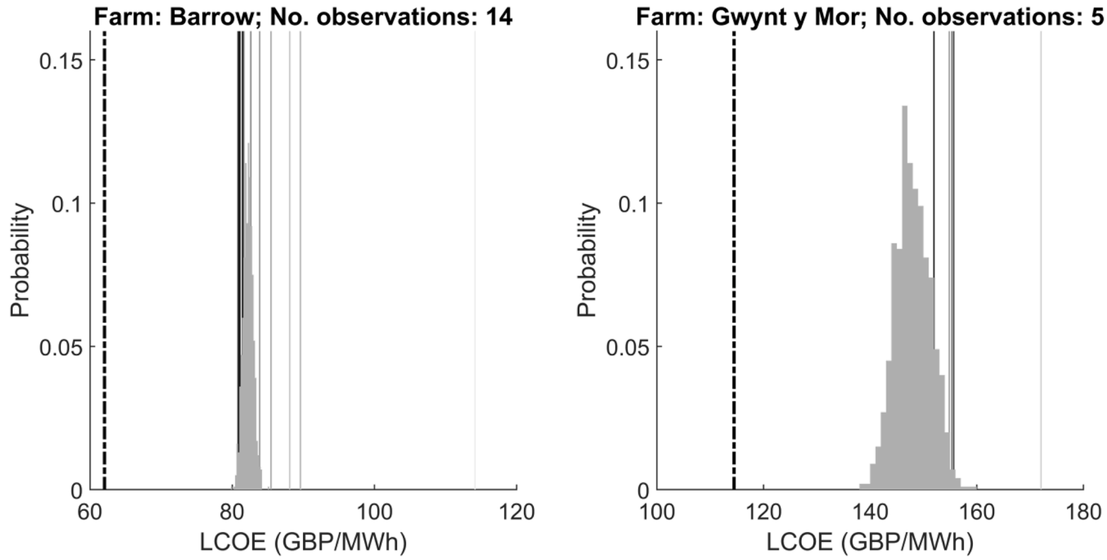

Fig. 11. Samples drawn from the posterior of the Bayesian model for the Barrow and Gwyn y Mor OWFs. Lines indicate the observed data for the OWFs, with the oldest observation being shown in grey and the newest in black. The dot-dashed line indicated the BEIS assumed capacity factor of 48%.

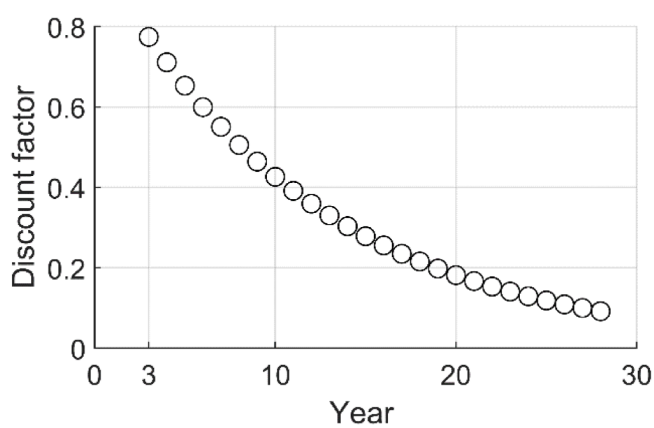

Fig. 12. Discount factor for a OWF with a 25 year life assuming a rate of 8.9%.

the more substantial Round 1 observations that form the prior. This gap will close over time as more data becomes available for Round 2 farms, as is the case in the Round 1 farms. In general, the Round 2 OWFs are found to have on average higher capacity factors (41.25% vs 35.18% using the yearly model) and lower standard deviations (3.60 vs 3.66) than the Round 1 farms. This is indicated in Table 2, where the average mean parameter for the Round 2 wind farms is larger. In addition, the average posterior resample of the observed data was higher for the Round 2 farms 1,000 times in 1,000 resamples. This justifies the higher capacity factors typically assumed for Round 2 farms, and the reduced standard deviation suggests less fluctuation in their generation.

In comparison to the values summarised in Section 1, all the values used in the literature to represent a typical Round 1 farm are within one standard deviation of the Round 1 mean capacity factor calculated by the proposed Bayesian model. However, the model mean is less than that used by Atkins and the BEIS, suggesting that their OWF capacity factors predictions are optimistic compared to the observed data.

3.3. Levelized cost of energy

The capacity factor influences the LCoE as described in Section 2.4. The samples predicted by the monthly and yearly model are broadly similar when converted into yearly capacity factors, as demonstrated in Fig. 9. Hence, this section is based on only the yearly model.

The LCoE was calculated based on the method described in Section 2.4. A number of different assumptions were implemented:

- LCoE calculated using the BEIS assumption of 48% capacity factors (dot-dash line).
- LCoE calculated using the mean of the capacity factor observations. A dotted line is used to show this in Fig. 11. The mean calculation was applied recursively using observations from the first year of operation to the current year. The line darkness is used to indicate this, with the darkest value for the LCoE calculated using all available capacity factor data. (See Fig. 12).
- Posterior predictions from the Bayesian model (shown as a histogram or summarised by important statistics in Table 3).

The results are shown in Table 3 for each OWF with accurate capital expenditure data, and also as histograms in Fig. 10 for Barrow, a Round 1 farm, and Gwynt y Mor (a Round 2 farm).

The average Round 2 wind farm has a higher LCoE value than the Round 1 OWFs. This observation is confirmed by Aldersey-Williams et al. [26]. The average of the Round 2 OWFs in their study is £20 higher than the Round 1 OWFs, while the LCoE values are £10 higher when averaged in the Bayesian model proposed in this study. This is a consequence of the larger capital expenditure for the Round 2 farms.

Also, when calculating LCoE values, it is observed that the Bayesian model balances between the observed data for the OWF being investigated, the others in the study and the diffuse prior. This is particularly useful for OWFs with a limited production history with a large standard deviation, such as West Duddon sands. However, a very low coefficient of variation is observed in the LCoE values for OWFs with a small standard deviation in their capacity factor. This is caused by discounting the annualised energy production in Eq. (11) and plotted in Fig. 11. Assuming a rate of 8.9% as in the analysis and taking the Barrow OWF as an example, the first year in which the Bayesian model is used to predict the AEP is year 17 (14 observations plus an assumed three years of construction), which is discounted by a factor 0.24. For West Duddon sands, the equivalent factor is 0.60; however, the lower standard deviation in the capacity factor acts to keep the variation in LCoE low.

3.3.1. Round 1 vs round 2 offshore wind farm

The LCoE was evaluated for a generic Round 1 and Round 2 OWF to assess the impact of random capacity factors. The yearly model was used to predict capacity factors. It had a Normal distribution with the

Table 3
Summary of LCoE values predicted for UK OWFs, all values are in £/MWh.

Farm	Round	BEIS assumption (capacity factor = 48%)	Mean - individual OWF	Posterior mean	Posterior standard deviation	Posterior 5%	Posterior 95%
Barrow	1	62.04	81.44	82.10	0.63	81.03	83.14
Burbo Bank	1	70.23	96.74	96.46	0.71	95.26	97.58
Inner Dowsing	1	80.07	105.87	106.72	1.01	105.14	108.48
Kentish Flats	1	54.66	80.94	81.49	0.58	80.53	82.44
Lynn	1	81.98	109.22	110.30	0.97	108.77	111.95
North Hoyle	1	72.69	101.11	101.33	0.62	100.33	102.41
Ormonde	1	103.29	126.02	127.33	1.55	124.79	129.91
Rhyl Flats	1	98.92	133.79	134.45	1.22	132.42	136.46
Robin Rigg (East)	1	97.55	130.05	130.12	1.10	128.34	131.90
Scroby Sands	1	84.17	124.30	126.97	0.86	125.56	128.37
Teesside	1	117.77	158.97	159.09	3.39	153.74	165.01
Burbo Bank Extension	2	131.97	158.75	158.05	2.28	154.26	161.88
Dudgeon	2	78.43	81.19	83.71	1.38	81.52	86.06
Greater Gabbard	2	109.03	124.89	125.30	1.24	123.29	127.40
Gwynt y Mor	2	114.49	151.98	148.78	2.25	145.16	152.66
Humber Gateway	2	116.40	130.59	131.84	1.81	128.81	134.60
Lincs	2	121.32	140.92	140.93	0.99	139.25	142.60
London Array	2	106.84	125.72	126.00	1.63	123.36	128.76
Sheringham Shoal	2	124.60	150.33	150.34	0.89	148.95	151.87
Thanet	2	107.11	150.45	149.69	1.67	146.89	152.37
Walney Phase I	2	100.28	119.13	119.04	0.82	117.76	120.38
West of Duddon Sands	2	62.31	66.86	67.16	0.44	66.42	67.86
Westermost Rough	2	91.82	96.42	100.44	2.11	97.05	104.14
Average		95.13	119.38	119.90	1.31	117.77	122.10
Average Round 1		83.94	113.50	114.22	1.15	112.36	116.15
Average Round 2		105.38	124.77	125.11	1.46	122.73	127.55

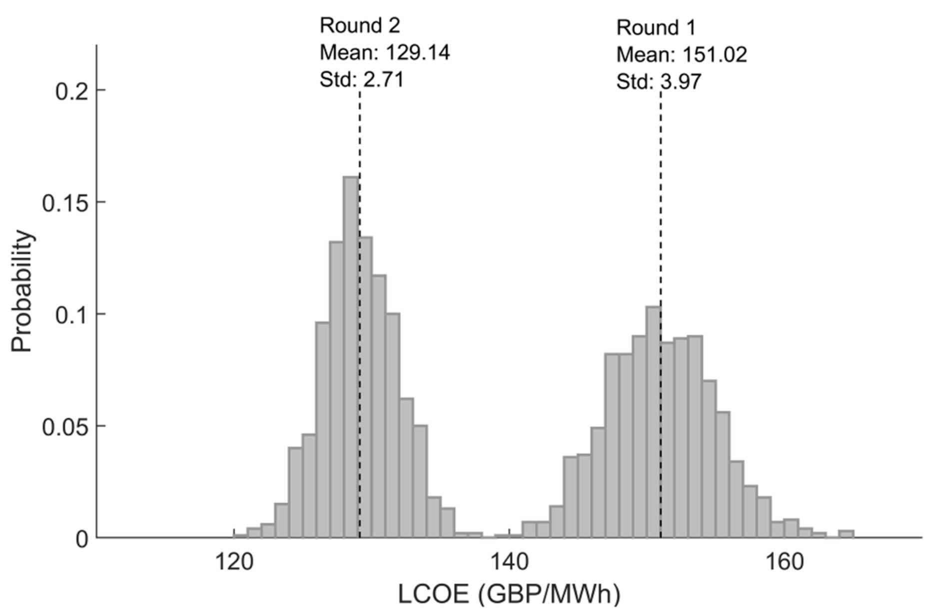

Fig. 13. LCoE for a typical Round 1 and 2 OWF.

parameters being the level 2 round mean, and the round mean standard deviations. In this calculation, cost data from a generic OWF defined by Ioannou et al [33] was used to allow comparison between the Round 1 and 2 predictions. The capital expenditure was taken as £1,551,720 k and the operational expenditure £56,597 k/year for a 504 MW capacity OWF. All other cost variables were the same as described in Section 2.2.

The results are shown in Fig. 13 and confirm that the Round 2 capacity factors result in lower LCoE than the Round 1 values for 1,000 Monte Carlo samples of the LCoE. Additionally, there is also a small reduction in the standard deviation due to the lower mean standard deviation in the Round 2 capacity factors.

4. Conclusions

This study proposed using a hierarchical Bayesian model to capture the variability in capacity factors yearly and monthly. The proposed Bayesian model can capture the large variability in OWF capacity factor predictions based on direct observations of the energy production. This has been demonstrated by developing both a monthly and yearly hierarchical model for OWFs in the UK. The monthly model additionally can capture variability throughout the year, where higher capacity factors are observed during the winter months. However, this model is more complex and converges slowly compared to the simpler yearly model.

Furthermore, it has more parameters and roughly ten times more observed data than the yearly model. Therefore, the proposed sampling algorithm requires substantially more computational effort. Additionally, when converted into a yearly capacity factor, it was found that the monthly model results in lower values for Round 2 farms than the yearly model (41.20% vs 35.33%). This is a consequence of the monthly model being more impacted by the farms for which there are many years of data available (e.g., the Round 1 farms). Updating the monthly model when new data is released will bring it into closer agreement with the yearly model.

The yearly model demonstrates that Round 2 wind farms indeed have higher capacity factors than Round 1 farms, as the distribution of averages is more than two standard deviations greater. In no case is a Round 1 wind farm observed to have a higher capacity factor than a Round 2 farm. Additionally, the average standard deviation in the yearly model of Round 2 farms was found to be smaller (3.60 vs 3.66). However, although there is a large variability in the capacity factor values, this was found to result in a smaller impact on the LCoE values and was demonstrated for a set UK OWF with accurate capital expenditure predictions. This is primarily because the capacity factor is discounted in future years of production, so even large variability in capacity factor has a diminished impact on LCoE. However, an accurate prediction of the capacity factor is still necessary for newer OWFs with less observed data, such as Gwyn y Mor, which had a standard deviation of more than double Barrow at 2.25 £/MWh.

The results from this study can be used both to predict capacity factors for individual wind farms or to make predictions for a generic wind farm (e.g., a typical Round 1 farm). Both are useful contributions because, as described in the literature review, simpler and potentially biased methods are currently used. The data is made available through Table 2, Table 3, or it can be extracted directly from the Bayesian model which is available through GitHub [24].

CRediT authorship contribution statement

David Wilkie: Conceptualization, Methodology, Software, Writing – original draft. **Carmine Galasso:** Supervision, Writing – review & editing.

Declaration of Competing Interest

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.

Acknowledgement

The authors appreciate the thoughtful comments of the Associate Editor and three anonymous reviewers that significantly improved the article. This research has been funded by an EPSRC Doctoral Prize Fellowship for UCL EP/R513143/1.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enconman.2021.114950.

References

[1] Walsh C. In: Offshore wind in Europe key trends and statistics; 2019. p. 2019.
[2] Brower MC. Wind resource assessment: a practical guide to developing a wind project. Wiley; 2015.
[3] Staffell I, Pfenninger S. Using bias-corrected reanalysis to simulate current and future wind power output. Energy 2016;114:1224–39. https://doi.org/10.1016/j.energy.2016.08.068.
[4] Bodini N, Optis M, Optis M. Operational-based annual energy production uncertainty: Are its components actually uncorrelated? Wind Energy Sci 2020;5:1435–48. https://doi.org/10.5194/wes-5-1435-2020.
[5] DNV GL. Potential to improve Load Factor of offshore wind farms in the UK to 2035 2019:1–83.
[6] Gonzalez E, Nanos EM, Seyr H, Valldecabres L, Yürüşen NY, Smolka U, et al. Key Performance Indicators for Wind Farm Operation and Maintenance. Energy Procedia 2017;137:559–70. https://doi.org/10.1016/j.egypro.2017.10.385.
[7] Atkins. Engineering net zero. 2015.
[8] Department for buisness energy & industrial strategy. Electricity generation costs. 2016.
[9] International Renewable Energy Agency. Renewable Power Generation Costs in 2018. Abu Dhabi: 2020.
[10] Danish Energy Agency. Finding your cheapest way to a low carbon future: The Danish levelized cost of energy calculator. 2020.
[11] National Renewable Energy Laboratory. Annual technology baseline. 2020.
[12] Altan A, Karasu S, Zio E. A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimiser. Appl Soft Comput 2021;100:106996. https://doi.org/10.1016/j.asoc.2020.106996.
[13] Karasu S, Altan A. Recognition model for solar radiation time series based on random forest with feature selection approach, n.d., p. 8–11.
[14] Department for buisness energy & industrial strategy. Digest of United Kingdom energy statistics 2019. 2019.
[15] Ayodele TR, Jimoh AA, Munda JL, Agee JT. Wind distribution and capacity factor estimation for wind turbines in the coastal region of South Africa. Energy Convers Manage 2012;64:614–25. https://doi.org/10.1016/j.enconman.2012.06.007.
[16] Crabtree CJ, Zappalá D, Hogg SI. Wind energy: UK experiences and offshore operational challenges. Proc Inst Mech Eng Part A J Power Energy 2015;229(7):727–46. https://doi.org/10.1177/0957650915597560.
[17] Aldersey-Williams John, Broadbent Ian D, Strachan Peter A. Analysis of United Kingdom offshore wind farm performance using public data: Improving the evidence base for policymaking. Util Policy 2020;62:100985. https://doi.org/10.1016/j.jup.2019.100985.
[18] Smith A. UK offshore wind capacity factors n.d. https://energynumbers.info/uk-offshore-wind-capacity-factors [accessed November 23, 2020].
[19] Sevieri G, De Falco A. Dynamic structural health monitoring for concrete gravity dams based on the Bayesian inference. J Civ Struct Heal Monit 2020;10(2):235–50. https://doi.org/10.1007/s13349-020-00380-w.
[20] Obane H. Forecasting photovoltaic and wind energy capital costs in Japan: A Bayesian approach. Energy Procedia 2019;158:3576–82. https://doi.org/10.1016/j.egypro.2019.01.908.
[21] 4COffshore. 4C Offshore - Global Offshore Wind Farm Database 2020. https://www.4coffshore.com/windfarms/ [accessed February 28, 2020].
[22] OfGem. the OfGem Renewables and CHP Register n.d. https://www.renewablesandchp.ofgem.gov.uk/ [accessed September 22, 2020].
[23] Ofgem. Renewable Energy Guarantees of Origin (REGOs): Guidance for generators , agents and suppliers. 2011.
[24] Wilkie D. BayesCapFacMonthly n.d. https://github.com/DavidOWT/ [accessed January 28, 2021].
[25] Elexon. The Elexon Portal n.d. https://www.elexonportal.co.uk/news/latest?cachebust=xlkzl1m6xd [accessed September 22, 2020].
[26] Aldersey-Williams J, Broadbent ID, Strachan PA. Better estimates of LCOE from audited accounts – A new methodology with examples from United Kingdom offshore wind and CCGT. Energy Policy 2019;128:25–35. https://doi.org/10.1016/j.enpol.2018.12.044.
[27] Bosch Jonathan, Staffell Iain, Hawkes Adam D. Global levelised cost of electricity from offshore wind. Energy 2019;189:116357. https://doi.org/10.1016/j.energy.2019.116357.
[28] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. 3rd ed. 2020.
[29] Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm. Bernoulli 2001;7:223–42. https://doi.org/10.2307/3318737.
[30] Marsaglia G. Evaluating the normal distribution function. J Stat Softw 2004;11:1–7.
[31] Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. J Comput Graph Stat 1998;7(4):434–55. https://doi.org/10.1080/10618600.1998.10474787.
[32] Short W, Packey D, Holt T. A manual for the economic evaluation of energy efficiency and renewable energy technologies. Renew Energy 1995;95:73–81. https://doi.org/NREL/TP-462-5173.
[33] Ioannou A, Angus A, Brennan F. A lifecycle techno-economic model of offshore wind energy for different entry and exit instances. Appl Energy 2018;221:406–24. https://doi.org/10.1016/j.apenergy.2018.03.143.