Energy Reports 9 (2023) 292–302

Contents lists available at ScienceDirect

Energy Reports

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

Research paper

Long-term wind and solar energy generation forecasts, and optimisation of Power Purchase Agreements

J.J. Mesa-Jiménez [a,b,*], A.L. Tzianoumis [c], L. Stokes [d], Q. Yang [a], V.N. Livina [b]

[a] Brunel University London, Kingston Lane, Uxbridge, UK
[b] National Physical Laboratory, Hampton Road, Teddington, UK
[c] City University of London, Northampton Square, London, UK
[d] Mace Group, 155 Moorgate, London, UK

ARTICLE INFO

Article history:
Received 28 July 2022
Received in revised form 18 October 2022
Accepted 26 November 2022
Available online xxxx

Keywords:
Power purchase agreements
Markov chain Monte Carlo
Stochastic forecast
Renewable energy optimisation

ABSTRACT

Due to more affordable solar and wind power, and the European Union regulations for decarbonisation of the economy, more than 40% of the Fortune 500 companies have targets related to green energy. This is one of the main reasons why multi-technology Power-Purchase Agreements (PPAs) are becoming increasingly important. However, there are risks associated with the uncertainty and variable generation patterns in wind speed and solar radiation. Moreover, there are challenges to predict intermittent wind and solar generation for the forecasting horizon required by PPAs, which is usually of several years. We propose a long-term wind and solar energy generation forecasts suitable for PPAs with cost optimisation in energy generation scenarios. We use Markov Chain Monte Carlo simulations with suitable models of wind and solar generation and optimise long-term energy contracts with purchase of renewable energy.

1. Introduction

Energy is the backbone of modern economy, with increasing need of electrification of various areas (such as transport) under the policies related to climate change, sustainability, and security of supply. Consumers, investors and politicians have been working on developing more efficient and sustainable ways to meet their electricity requirements.

According to Fulbright et al. (2021), more than 40% of Fortune 500 companies have targets related to renewable energy procurement, energy efficiency or cutting Greenhouse Gas Emissions (GHG). One example is Google, whose target in 2010 was to achieve 100% renewable energy by 2017 for their global operations, including data centres and offices (Hölzle, 2017).

With development of more efficient solar power technologies, this type of renewable energy supply becomes a viable option, economically and environmentally, for development of energy-demanding industries, such as crypto-currency mining (Nikzad and Mehregan, 2022) and field irrigation (Nikzad et al., 2019). Tesla is building a solar farm of 3.8MW for bitcoin mining (More, 2022) in Texas, US. This illustrates the new technologies appearing on the interface of renewable energy generation and user demand needs.

As defined in Jenkins et al. (1999), a *Power Purchase Agreement* is a long-term contract between an Independent Power Producer (IPP) and an off-taker, usually an energy-intensive organisations or an utility company. PPAs are seen as a hedging tool by many organisations, as they offer an opportunity for energy buyers to achieve price certainty beyond 3–5 years, and at the same time meet their sustainability objectives.

Mature renewable technologies were price-competitive in 2020 and could offer prices for their intermittent output at all-time lows. In 2021, economic turmoil has affected most commodity markets, and as a result, Levelized Costs of Electricity (LCOEs) of all electricity-producing technologies. Renewables in 2022, still offer electricity at a discount compared with fossil-fuelled power plants. This trend, albeit being delayed by various economic circumstances, defines the future energy systems.

The intermittency of renewable technologies is one of the main challenges in renewable PPAs. Many corporations are reluctant to be exposed to this risk, despite the fact that it constitutes only a small fraction of the total value of a PPA.

In physically-settled PPAs, which is the most commonly used structure, there is usually a Balancing Responsible Party that undertakes the balancing tasks, which are necessary to achieve effective hedging via PPAs. The balancing responsibilities should

* Corresponding author at: National Physical Laboratory, Hampton Road, Teddington, UK.
E-mail addresses: joaquinmesajimenez@gmail.com (J.J. Mesa-Jiménez), tzianoumis@gmail.com (A.L. Tzianoumis), Lee.Stokes@macegroup.com (L. Stokes), qingping.yang@brunel.ac.uk (Q. Yang), valerie.livina@npl.co.uk (V.N. Livina).

https://doi.org/10.1016/j.egyr.2022.11.175

reflect three types of risks, which are related to variable generation patterns (Hedges et al., 2019):

- **Balancing risk**: Risk associated to the exposure of power system costs that arise when an asset's forecasted generation is different from its actual generation. This risk is related to the imbalance cost, therefore the more an asset contributes to the power system's imbalance, the higher the cost.
- **Shape or profile risk**: This risk is related to the variability of wind speed or solar irradiation, and is independent of the total volume generated by the asset, which will differ from a 24-hour base-load delivery quoted for standard products.
- **Volume risk**: It captures the variable generation of an asset over a certain period of time. This can be related to deviations in the long-term, such as higher than expected wind speed or lower levels of irradiation due to, for example, abnormal weather conditions or unplanned outages.

Risks can be mitigated depending on the structure of the PPA contract itself (Brindley, 2020), but also by achieving the best possible forecast for both wind and solar generation. We consider a long-term scale forecasting horizon, of at least one year, for wind and solar generation.

Wind generation forecasting has always been of interest for energy community, as estimation of wind generation forecast influences sizing reserves and shape-balancing risk (Gil et al., 2010; Constantinescu et al., 2009; Lowery and O'Malley, 2012; Mauch et al., 2013). The authors (Wilczak et al., 2015) point out the importance of wind forecast, but their aim is to improve the accuracy of short-term wind forecast. In another example, Solari et al. (2012) uses geostrophic wind data to forecast wind speed for port safety, but the maximum horizon achieved is in the order of days. Similarly, Cheng et al. (2017) aims to forecast wind speed using anemometer data, but the forecast horizon is short-term. Also, Lange et al. (2006) outlines several techniques for wind forecast, but it follows a similar forecast horizon. Many wind forecast horizons are oriented to achieve short-term accurate results (Huang et al., 2011; Akçay and Filik, 2017; Bossanyi, 1985), even though what is considered in this field as long-term does not extend to more than 120 h, as is the case of Barbounis and Theocharis (2006). A case of wind speed long-term forecast is presented in Azad et al. (2014), where a series of Neural Network (NN) methodologies are used to forecast wind speed in a period of six months by using only previous wind speed pattern data. Other predictions are short-term (Negnevitsky and Potter, 2006; Shi et al., 2013; Pinson et al., 2009; Juban et al., 2007).

We are interested in estimating electricity generation, in particular, in the probability of generation values to happen within several intervals during the year than in a forecast for the next period. Matos and Bessa (2010), Pinson et al. (2007) use probabilistic modelling and Markov Chain Monte Carlo (MCMC) process to generate the wind field data, and they do so for a time horizon of maximum 50 h. There are many studies based on Bayesian methodologies to model short-term wind speed forecast (Jiang et al., 2013; Bracale and De Falco, 2015; McLean et al., 2013). As mentioned before, many models use meteorological data to forecast energy generation, but the inconvenience of this is that the forecast goes as far as the meteorological model goes, and this is usually short-term.

Concerning solar forecast, the challenge is minor in comparison to wind forecast. This is due to the hours of daylight which are known. There are other climatological factors that influence solar generation, such as cloudiness. Solar radiation is commonly forecast in order to estimate solar energy generation (Reikard, 2009; Heinemann et al., 2006; Sfetsos and Coonick, 2000; Perez et al., 2010). One of the most popular approaches is NN-based models, such as Gensler et al. (2016), which uses autoencoders and Long Short-Term Memory (LSTM) NNs to perform the forecast, see also (Chen et al., 2011; Abuella and Chowdhury, 2015). When looking at the forecasting horizon, similar to wind, most of the predictions are focused in the short term (Urquhart et al., 2015; Golestaneh et al., 2016; Bacher et al., 2009), therefore they cannot be used for the purpose of the paper.

The goal of this paper is to produce long-term forecasts of wind and solar energy generation combined, for the purposes of PPAs, with time horizon of one year, taking into account the three types of risks in order to find an optimal match of the forecasts with respect to the target consumption profile. The rest of the paper is organised as follows: In Section 2 we describe the problem in more detail and the data, as well as the characteristics and electricity prices of the considered wind and solar farms. In Section 3 we outline the methodology used to achieve the goal of the paper for solar and wind energy, and the optimisation problem. In Section 4 we explain the obtained results. Finally, limitations and further work are outlined in Section 5, and conclusions in Section 6.

2. Problem description & data

The objective of this work is two-fold: estimation of solar and wind generation for the next year (and further few years if necessary), and optimisation of price profile and shape for the estimated solar and wind profiles with respect to a determined business consumption profile. The estimation of both wind and solar generation relies only on observed data, therefore no other predictors are used. We estimate the electricity prices by averaging a certain range of prices in the year 2018. Information about total volume generated and prices for 2018 can be seen in Table 1:

Temporal resolution of data is half-hourly, therefore the total length of the data is 17520 points for all energy farms, and for the consumption profiles. Specific characteristics, such as turbine model and number or farm locations, are not provided, due to data confidentiality.

3. Methodology

In this section we outline the methodologies used for wind and solar energy generation estimations, and the optimisation problem. Due to the difference between wind and solar, two different techniques have been used for each of them. We start by explaining Bayesian inference for wind, and a different algorithm for solar, both relying on one year data available for the training stage.

3.1. Wind generation estimation

3.1.1. Bayesian modelling and Markov chain Monte Carlo

Bayes theorem

Bayesian inference is the process of deducing properties of probability distribution from the observed data by using the Bayes theorem (Bayes, 1763). First, we define a prior distribution, by estimating the histogram of the observed dataset. We use the prior distribution as our belief, then the prior's parameters are updated according to the Bayes theorem:

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)} \tag{1}$$

Here, $P(A)$ refers to the prior distribution, $P(X|A)$ is the likelihood distribution, and $P(A|X)$ refers to the updated posterior probability.

Markov Chain Monte Carlo

According to Walsh (2004), the main idea of MCMC process is to have an estimate state $\mathbf{x}$ that begins as an arbitrary value

Table 1
Annual volume (MWh) and price (£/MWh) for different assets and consumption profile.

	Wind farm 1	Wind farm 2	Wind farm 3	Wind farm 4	Solar farm 1	Consumption profile
Annual volume (MWh)	78000	19500	55500	42000	59000	195000
Price (£/MWh)	51.50	53.98	55.60	61.03	48.40	

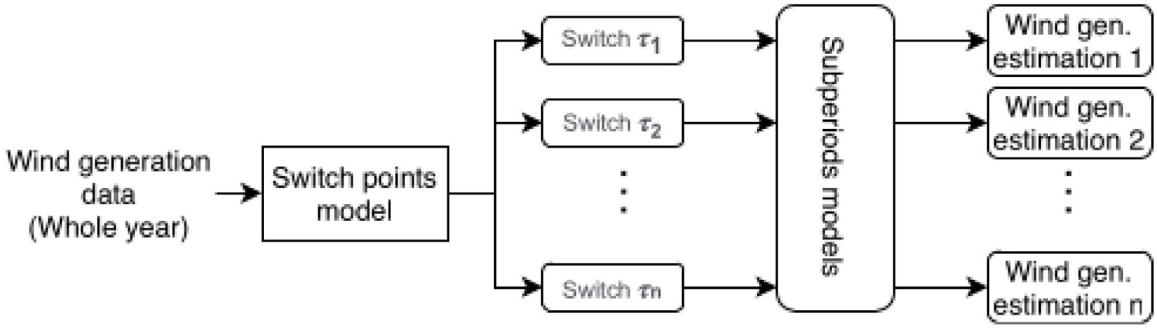

Fig. 1. Overview of the wind generation model. One-year data is used to estimate points where changes of pattern happen (switch points). A different model is then applied separately to every period of data to get estimates for the period.

and develops further with time steps. Eventually **x** becomes a fair sample from a certain objective distribution $p(\mathbf{x})$. Markov chains are sampled from some distribution $q^{(t)}(x)$, where t denotes the number of time steps occurred. At the beginning, the sampling algorithm departs from some distribution $q^{(0)}$ that arbitrarily initialises x for each chain. Then $q^{(t)}$ is influenced by all previous Markov chain steps. The goal for $q^{(t)}(x)$ is to eventually converge to $p(x)$.

Following Goodfellow et al. (2016), when a single Markov chain state x is updated to a new state x', the probability of a single state transitioning to state x' is given by

$$q^{(t+1)}(x') = \sum_x q^{(t)}(x) T(x'|x), \tag{2}$$

where $T(x'|x)$ is the transition distribution that specifies the probability that a random update will go to state x' when departing from state x. T can be re-written using matrix **A**, defined such that $A_{i,j} = T(\mathbf{x}' = i|\mathbf{x} = j)$, so the expression can be re-defined as

$$\mathbf{v}^{(t)} = \mathbf{A}\mathbf{v}^{(t-1)}. \tag{3}$$

3.1.2. Wind generation model

For the purpose of wind generation estimation, we use Bayesian modelling. It is known that wind patterns are very difficult to predict, therefore this methodology needs to handle some uncertainty. We rely solely on historical wind generation data to estimate wind generation. The fact of not using any climatological information adds an extra difficulty to the problem, as patterns and behaviour have to be inferred from historical data.

The first step we take to infer patterns in wind generation data is to detect changes on the wind generation distribution. Then, for every separation or period, we fit the data in each subset individually using a mixture model. The high level overview of the process is shown in Fig. 1.

We first assume that for every period, the wind energy generation patterns have a particular probability distribution. Then the model assigns the switchpoints where changes in the parameters of the probability distribution occur. Switch points are then included in a discrete Poisson distribution, whose parameters are defined according to the switch function implemented in the python library PyMC3 (Salvatier et al., 2016).

The schematic of the switch point detection model can be seen in Fig. 2. Here, the priors for the different switch points (or τ_s) have been defined with uniform distribution. We assume that wind generation data fits a beta distribution. The reason to select beta distribution is that it is very versatile: it can take 'U'-shape or just decrease from the left or increase towards the right side of the histogram. Beta distribution supports values between 0 and 1, therefore wind generation data has had to be rescaled within that range.

Once the data has been trained and the switch points have been obtained, we separate the data into different subsets and train every period individually. As mentioned in Section 2, we use the year of 2018 as the training period, and the timeline with the parameters is shown in Fig. 3.

For every subperiod, we use a mixture model composed of three beta distributions, whose parameters α and β are defined by a uniform distribution with values between 0 and 10. The weights of each distribution are defined by a Dirichlet distribution with $\alpha = 1$, under the condition that the sum of all of them is equal to 1.

The wind generation histogram accumulates more values near the minimum and the maximum, as observed in Fig. 4, where switch points are applied. It also may accumulate points around the centre of the histogram. This is the reason for choosing more than one distribution as a prior.

As can be seen in Fig. 4, wind generation patterns oscillate between the maximum and the minimum, with a significant amount of values on both extremes. This pattern is more obvious during winter and spring, as the values of energy generation shift to the left (closer to minimum generation) in summer. After the summer period, the pattern starts inclining again towards maximum generation values, being the maximum generated during the last period of the year.

3.2. Solar generation estimation

For the purpose of solar generation estimation, we develop a procedure that relies on historic (previous year) generation data, in order to extrapolate the generation to the following years to accommodate yearly decay of PV panels. The annual solar panel degradation can be assumed as 1% (Stahley, 2019). As the model does not simulate meteorological conditions, we need to add a

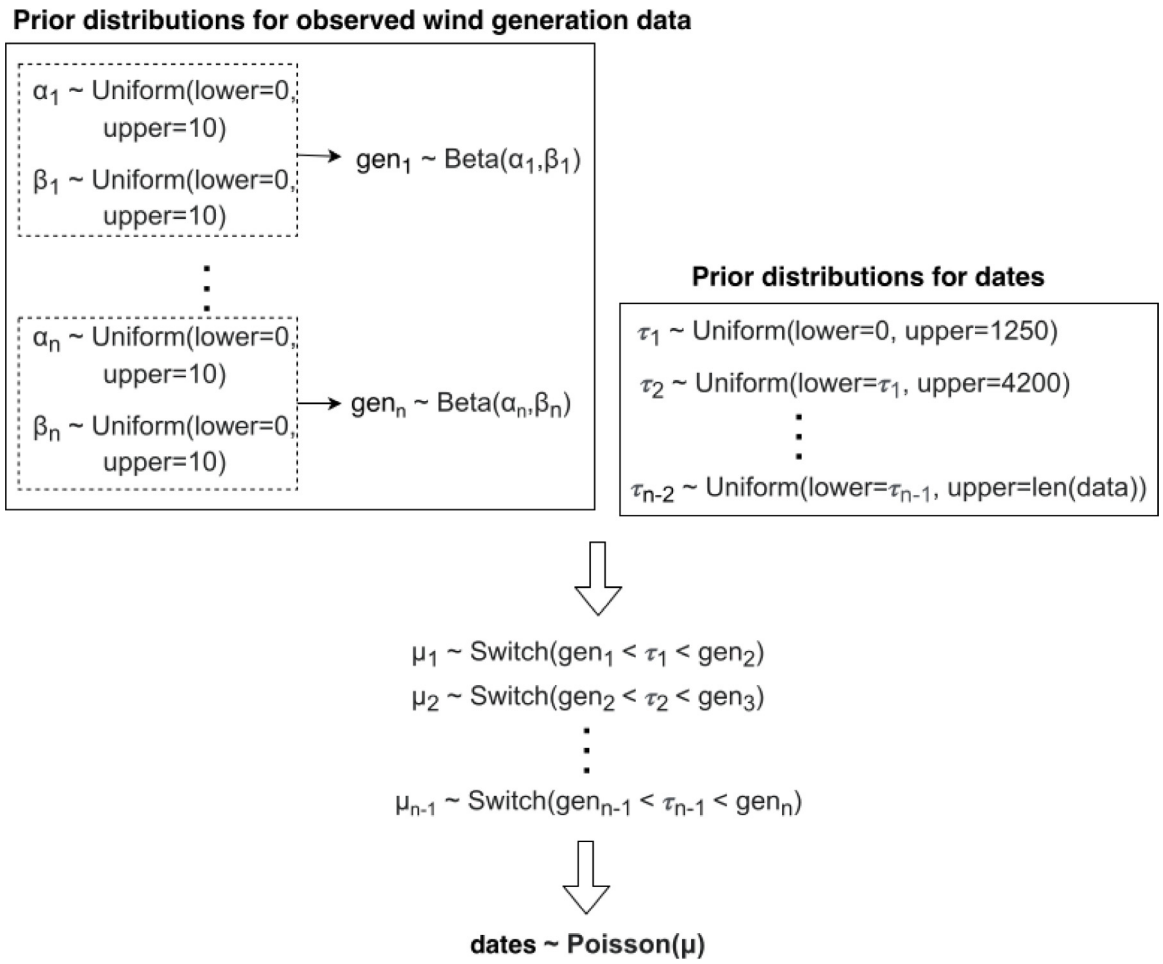

Fig. 2. Switchpoint model description. Priors for the beta distributions are defined by an uniform distribution between 0 to 10 for α and β. Parameters τ_s are defined by an uniform distribution too, where the lower value is the previous τ distribution and the upper one is restricted up to two months with respect to the previous upper limit, approximately. *gen* are generation subsets, which are optimal after model is trained with observed data.

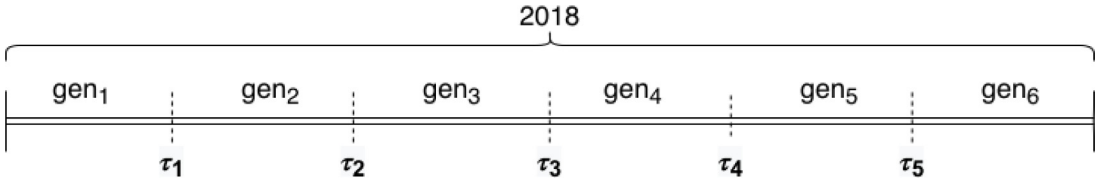

Fig. 3. Scheme for timeline defining the estimation of wind generation for year 2018. Six intervals are defined by five switchpoints.

noise component to approximate the stochastic variability in the observed data. For this purpose, we generate red noise (Gilman et al., 1963), with a scaling exponent $\beta = 0.9$, mean $= 1$, std $=$ 0.5.

The algorithm starts by extracting the rolling average of 30 half-hour periods. Next, the minimum and maximum of such rolling averages are extracted. Then, the lengths of the vectors for the numbers and hours of light are defined, as we need to define two main trends: seasonal and diurnal. For both, we use a sinusoidal wave. We combine diurnal and seasonal variability, with the maximum generation peaking at midday. Red noise vector of the same length as the training data is added as a multiplicative component, point-wise, to the obtained vector. Rate of decay or panel degradation is the final additive component, with the number of years assumed according to the forecast horizon.

Let $\boldsymbol{S_{roll}} = [S_{n-1}, \ldots, S_T]$ be the $n = 30$ period rolling average vector of solar generation, and $S_{\min}$, $S_{\max}$ the minimum and maximum values of $\boldsymbol{S_{roll}}$ where:

$$\overline{S}_t = \overline{S}_{t-1} + \frac{1}{n}\left(S_t - S_{t-(n-1)}\right), \text{ with } \overline{S}_t = \frac{1}{n}\sum_{i=0}^{n-1} S_{t-i}. \tag{4}$$

Let $\boldsymbol{t_s} = [0, \ldots, \pi]$ be a vector of length = *Number of days in the year* and $\boldsymbol{t_d} = [0, \ldots, \pi]$ be a vector of length = *Number of half-hour periods in a day*. We define M as the number of points in our training data. Therefore, we define the seasonal component of the model as:

$$y(t) = \sin(t) S_{\max}, \text{ with } t \in \boldsymbol{t_s} = [0, \ldots, \pi] \tag{5}$$

We define the seasonal component as follows. The value is set to the minimum of the season, S_{min}, when $y(t)$ is minor than this, therefore the result is:

$$f_{seasonal}(t) = \begin{cases} S_{\min} & \text{if } y(t) \leq S_{\min} \\ y(t) & \text{if } y(t) > S_{\min} \end{cases}$$

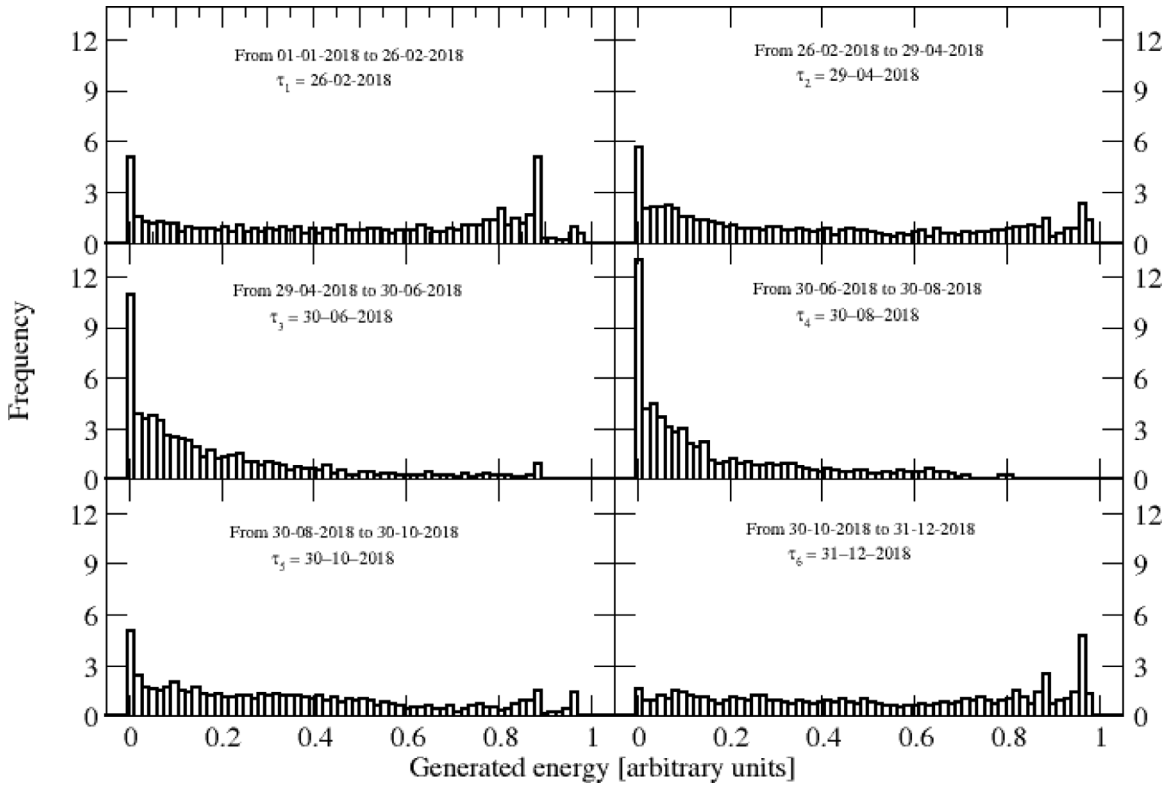

Fig. 4. Histograms of wind generation of the wind farm 1 in time intervals separated by switchpoints. The data in *x*-axis are rescaled into interval [0, 1] for comparison, so the units of energy are arbitrary.

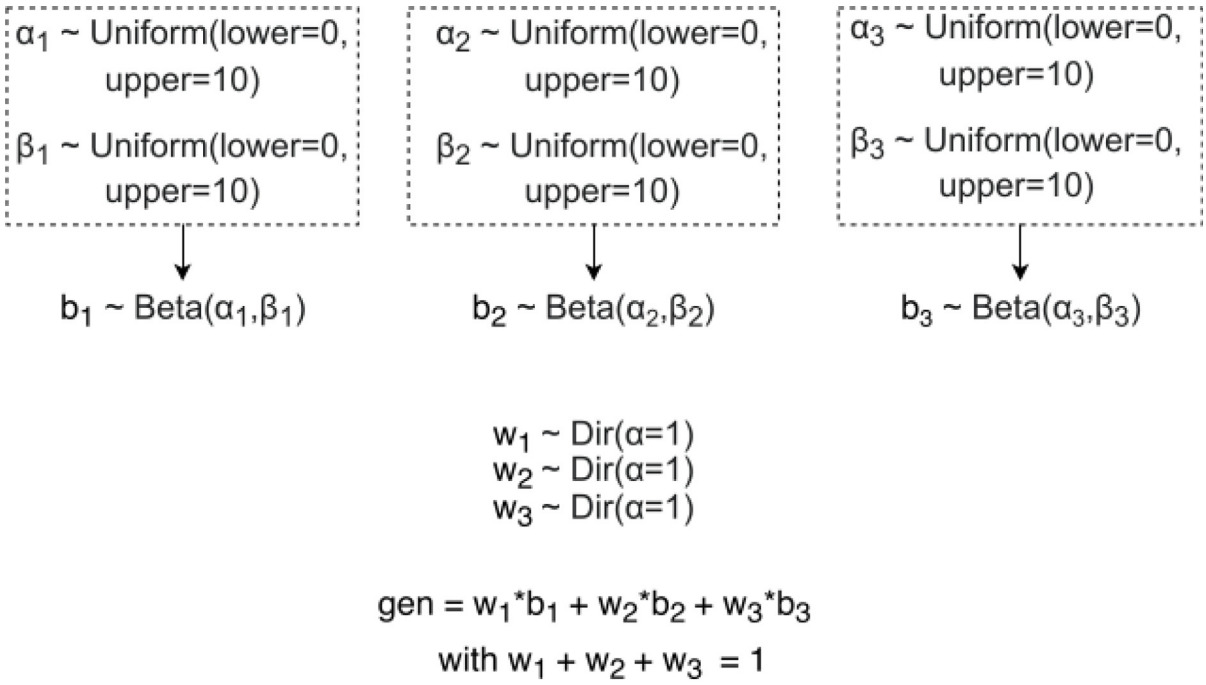

Fig. 5. Mixture model for every subperiod of a year. Such a model is formed of three beta distributions, whose priors are defined by a uniform distribution over intervals [0, 10] for both α and β. *Dir* is the Dirichlet distribution used to fit the weights.

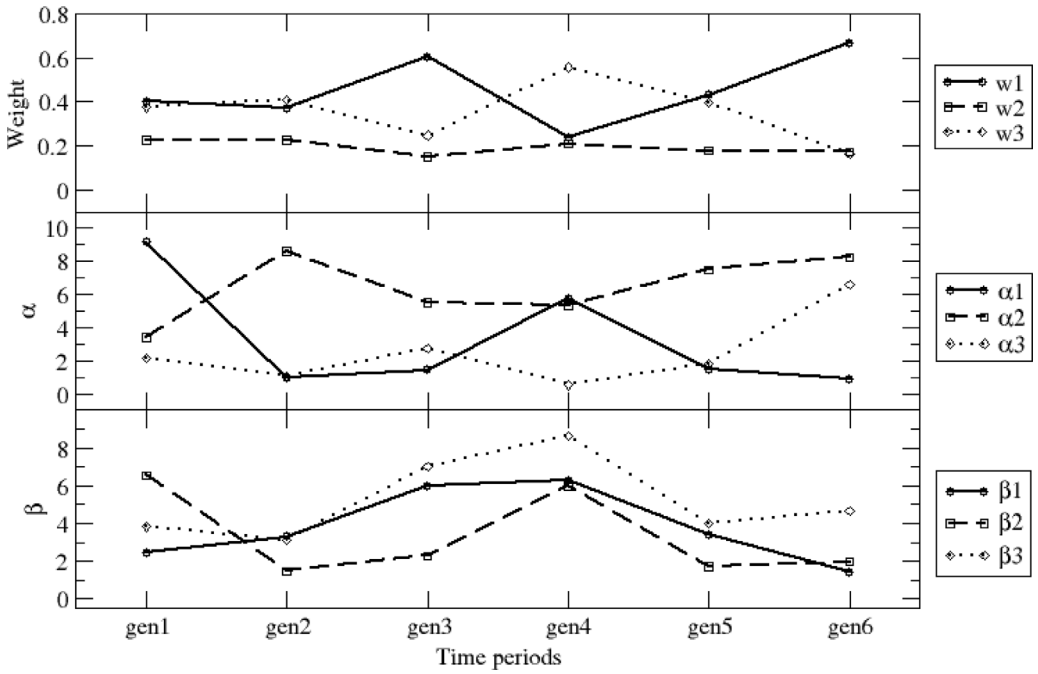

Fig. 6. Evolution of parameters for the Bayesian mixture model in each interval between switchpoints.

We add red noise η and rate of decay of 1% denoted as *decay*. Finally, the model equation of solar energy generation is as follows:

$$Solar(k) = (\mathbf{z}(k) + f_{seasonal}(t)\sin(\mathbf{t}_d)) \cdot decay(i) \cdot \eta,$$
$$\text{for } 24 < k + 48 < M;\ 0 < t < \text{number of days in a year};\quad (6)$$
$$0 < i < M$$

Using this model, we formulate the optimisation problem.

3.3. Definition of the optimisation problem

We prioritise the price of newly build assets over the demand profile in the optimisation problem, as in a real case scenario a client would seek to fulfil electricity demand by paying the minimum price possible. In a real-case scenario there are other less quantifiable parameters that would influence buyer's decision, but in this paper we focus on the key parameter, which is the price of the commodity element of the Power Purchase Agreement.

$$\begin{aligned}
\text{minimise}\quad & \sum_{m=1}^{M} Ps_m \cdot \vec{S}_m + \sum_{k=1}^{K} Pw_k \cdot \vec{W}_k, \\
\text{subject to}\quad & \sum_{m=1}^{M} \beta_m \cdot \vec{S}_m + \sum_{k=1}^{K} \gamma_k \cdot \vec{W}_k \geq \vec{C}, \\
& \sum_{m=1}^{M} \beta_m + \sum_{k=1}^{K} \gamma_k \geq 1, \\
& m = 1, \ldots, M;\ k = 1, \ldots, K, \\
& \\
& \beta_b = 0, 1;\ \forall\, b \in 1, \ldots, M, \\
& \gamma_c = 0, 1;\ \forall\, c \in 1, \ldots, K,
\end{aligned} \quad (7)$$

where Ps and Pw are MWh prices for every solar and wind generation asset, $\vec{S}$ and $\vec{W}$ are the vectors containing generation data for solar and wind, respectively, $\vec{C}$ is the target consumption profile, and β and γ are the binary decision variables for each generation source (*'on'* or *'off'*). Therefore, this is a linear programming minimisation problem, whose goal is to minimise the price of generation for both assets, subject to fulfil the difference between the aggregated generation sources and the target consumption profile, in order to find the optimal combination. This optimisation problem has been implemented in Python using the *optimise* method from *SciPy v1.4.1* library (Jones et al., 2001).

4. Results

4.1. Results of the wind generation modelling

Our goal is to identify suitable switchpoints for data subsets and generate realistic data patterns using Bayesian modelling and MCMC for every period. Fig. 6 shows the weights and the parameter values α and β for the three distributions comprising the mixture model in each of the six intervals of the year. The weights show the importance of each of the distributions in each period of generation. The results are used for wind simulation in the following year, as our main interest is to capture a range where the wind generation is more likely to be homogeneous. The histograms of the simulated data for every period can be seen in Fig. 7.

In Fig. 7, one can see that the twenty simulations of the mixture model with the three beta distributions approximate the histogram of the data very well for every period. This figure displays the differences between summer and winter period in both actual and simulated data. Due to the smoothness of the mixture data, it can be seen that the simulated data could miss some spikes in generation data close to 0.9 in the right side, but due to the nature of the beta distribution, it captures the left side well in all cases. This shows the flexibility of beta distribution in all different scenarios.

We compare the generated simulations with their respective actual values observed in 2018. This is a hindcast exercise, i.e. a forecast in the past which can be compared with ground truth. Results can be seen in Table 2.

4.2. Results of the solar generation modelling

Fig. 8 shows actual solar generation for 2019 and its corresponding simulations for the same year. The reference year used

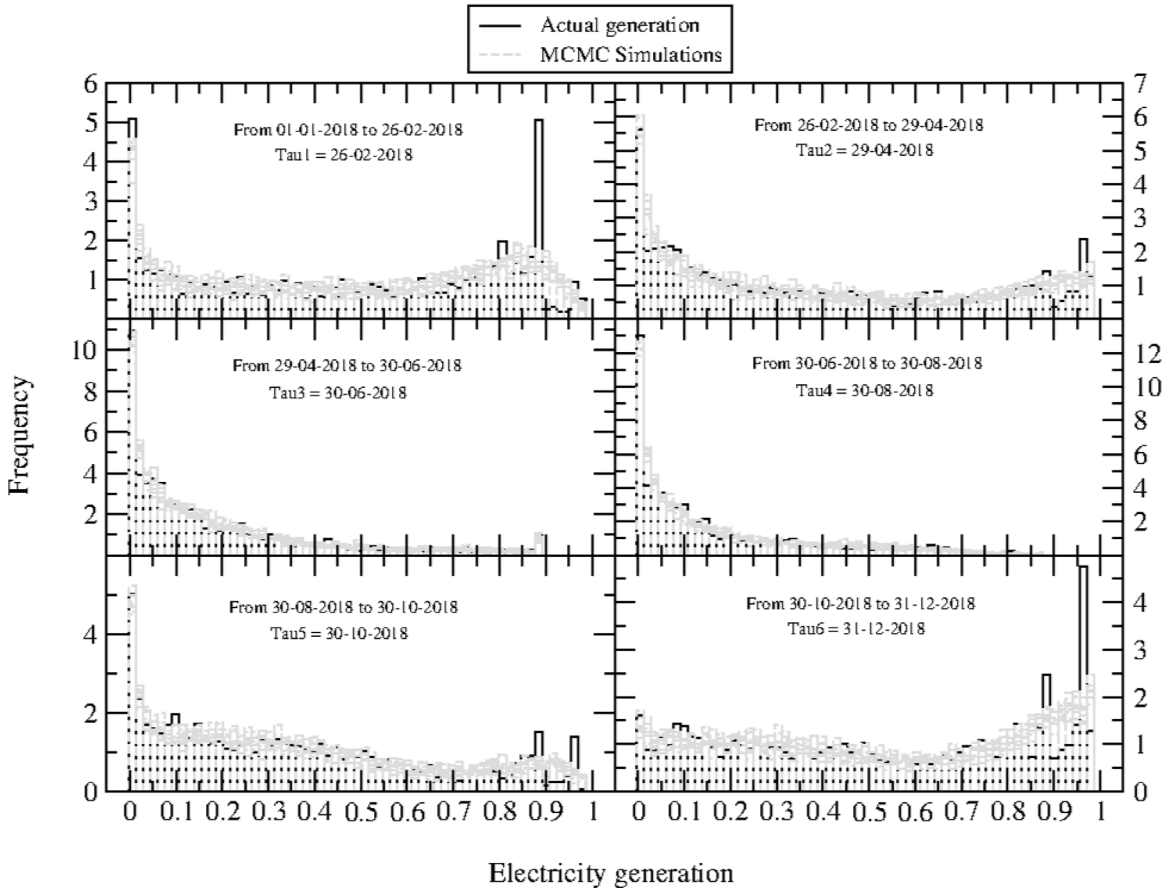

Fig. 7. Histograms of wind generation and MCMC estimations for wind asset 1 for separated subsets. A total of twenty MCMC traces have been used to create the histograms in every subperiod. The data for histograms was scaled between 0 and 1 for convenience of comparison.

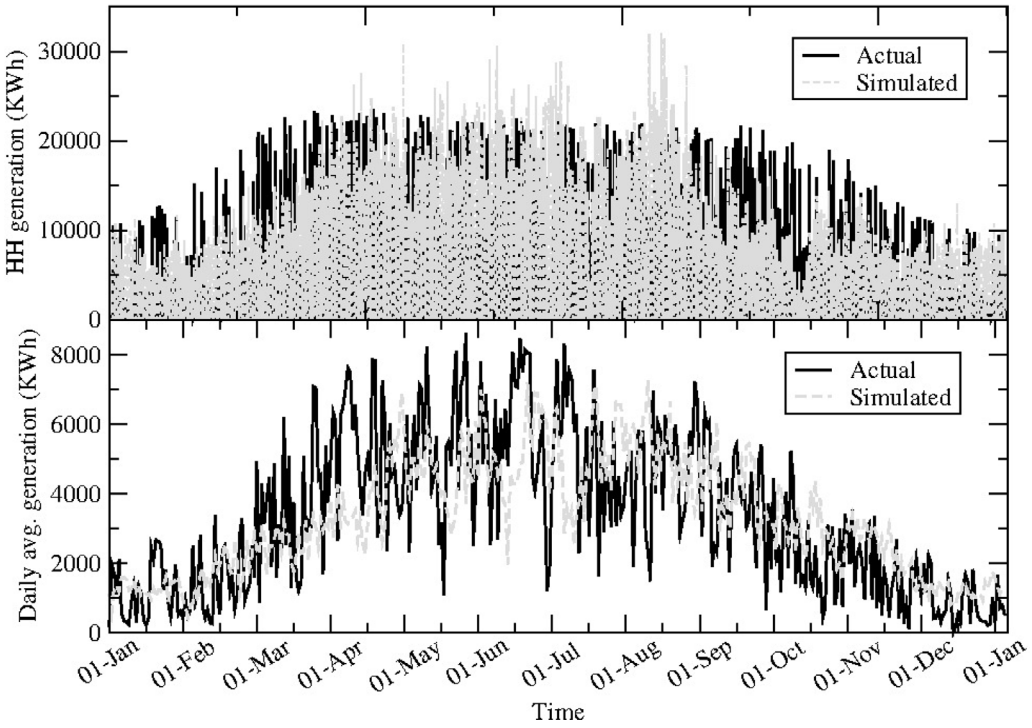

Fig. 8. Actual and simulated solar generation for the year 2019. The simulations have been carried out with the stochastic model described in Section 3.2. The top panel represents data of half-hourly temporal resolution. The bottom panel represents daily average generation. Year 2018 has been used as a reference period in order to simulate values to the following year.

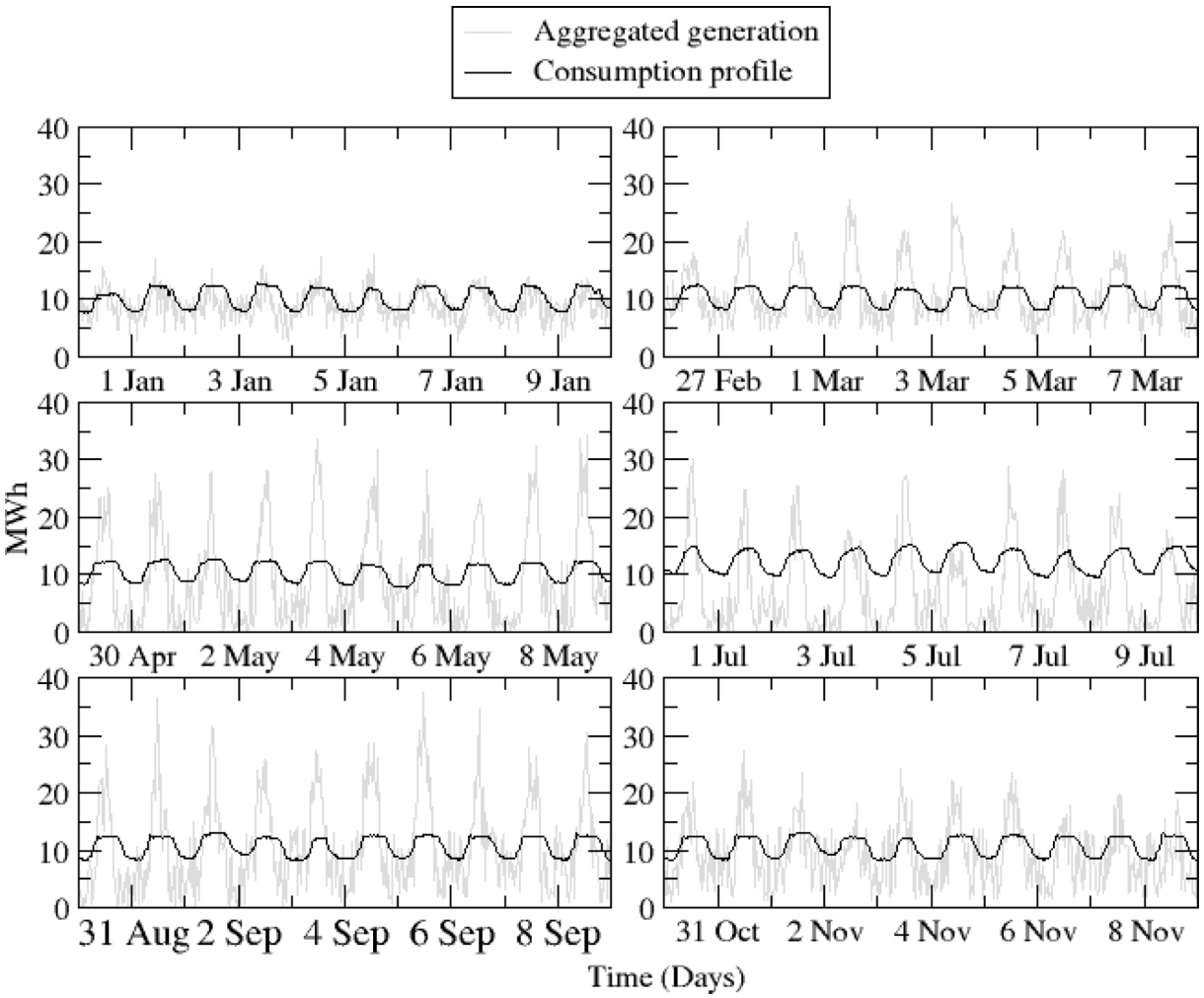

Fig. 9. Optimal mixed generation of wind and solar energy, and energy consumption profiles in 2019. Every panel corresponds to ten days in six separated periods of the year. Time resolution is half-hourly.

Table 2
Comparison of actual and simulated total annual volume for 2019 for the four wind farms (hindcast exercise).

	Wind farm 1	Wind farm 2	Wind farm 3	Wind farm 4
Actual volume (MWh)	77728.30	19579.12	55530.87	41753.34
Simulated volume (MWh)	80448.22	23164.32	60067.23	48774.21

Table 3
Comparison of actual and forecast total annual volume of solar generation for three consecutive years (three-year forecast horizon for 2018–2020), based on the data of 2017.

	2018	2019	2020
Actual volume (MWh)	58776.61	62362.27	58424.66
Simulated volume (MWh)	59845.69	59789.51	59955.11

is 2018, therefore we applied rate of decay of PV energy generation over one following year. As the model considers minimum and maximum peaks on the range of its moving average with period of 30 half-hourly units, the path followed by the seasonal line is similar. Fluctuations within the same day are modelled by the multiplicative red noise, as described earlier.

We show daily average values displaying the seasonal trends and the noise. The noise amplitude increases from winter to summer, and then decreases back to winter levels. These irregularities are followed by the simulated data, which shows not only the same seasonality pattern, but also the same noise pattern.

We also compared the forecast for three years, in terms of the total volume in MWh: we considered data of 2017 as historic and forecast ahead for the three years 2018, 2019, 2020, with a longer forecast horizon than in the previous experiments. These results can be seen in Table 3.

Now that the estimations for both wind and solar generation have been obtained for wind asset 1 and solar asset, we apply the same model to the rest of assets and find the best combination that matches them and minimises total cost and difference of shape between total aggregated and consumption.

4.3. Linear programming for optimal shape match

Before proceeding with finding the optimal combination of wind and solar generation, we perform the estimation of generation of all assets. For the consumption profile, we assume that a very similar pattern repeats year after year, as we have observed based in the historical data.

As shown in Fig. 9, total aggregated generation sometimes exceeds the consumption curve. It struggles to reach consumption

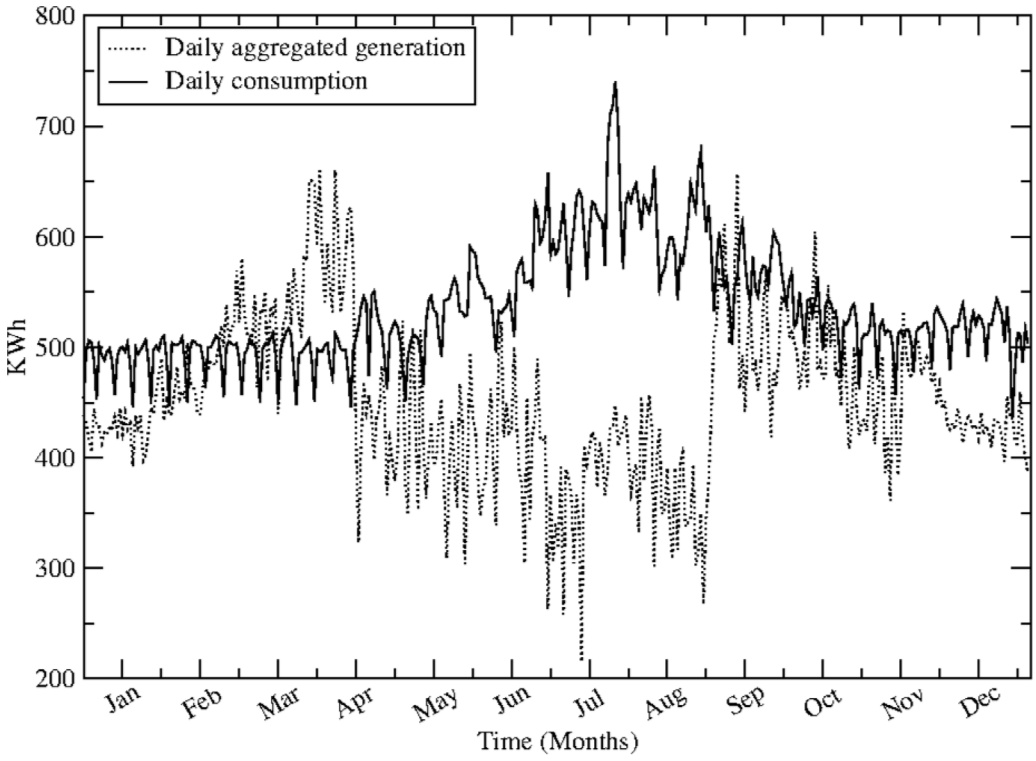

Fig. 10. Optimal daily mixed generation of wind and solar energy and energy consumption in 2019.

at the baseload during the summer period, due to the lack of wind generation available. However, there is normally a compensation of solar generation within the day during peakload.

Fig. 10 shows daily aggregated data for both optimal generation and consumption, where the lack of wind generation in summer can be seen. Also, the plot shows that the pattern of the consumption profile increases in summer (most likely, due to air conditioning), therefore the difference would be even higher. Wind generation tends to decrease in summer, yet solar generation is higher. The gap found in summer is due to two reasons: a predominant number of wind generation assets over solar assets, and the seasonality of the consumption profile.

Optimisation has been performed over the half-hourly data (Fig. 9), therefore the optimised solution accounts for every half-hourly model estimates to be as close as possible in total. If the optimisation was performed in the total aggregated daily data, then the optimisation would have been oriented to this specific resolution and the results would have been slightly different. We chose to optimise half-hourly resolution data to assess more accurately the amount of electricity to buy or sell within the day. This helps one to assess the risk of purchasing energy blocks in the intraday or mid-term energy market beforehand.

5. Limitations and further work

The main limitation is related to data availability and quality. For the purpose of this paper, we obtained two-year data for every wind farm, and three years for the solar farm. Ideally, we would like to test this model by using a longer forecasting horizon, especially for wind. Also, regarding data availability, we could only get data from one solar farm, so repeating the experiment with a higher number of solar farms would give a wider perspective of solar energy market.

As we see in the results for wind simulations, we achieved satisfactory results by breaking the problem down to six subintervals of a year using five switchpoints. However for further research on this area, it would be interesting to create further divisions in order to consider individual estimations and finding an optimal number of switchpoints, preventing overfitting.

6. Conclusions

In this paper, we applied two techniques, Bayesian estimates and MCMC processes, to model wind energy generation: one for detecting change points in the data using beta distribution, and another for period separation. We proposed a stochastic model to forecast solar generation.

MCMC is a flexible and reliable method that suits the approach used in the paper. Various methods can be used to model stochastic systems, and in a recent paper (Billuroglu and Livina, 2022), a comparison of ML and ARIMA models was performed, which was the goal of the exercise. In this paper, our goal is to demonstrate the application of PPA with stochastic modelling, and MCMC adequately addresses the need.

In order to find the optimal combination of the results of the forecasting and estimation with respect to a particular consumption profile, we used linear programming to minimise costs and fulfil the required electricity demand. The optimal combination of intermittent output of renewable projects, could support energy buyers to minimise the commodity element of the PPA, the related balancing costs on the PPA and, last but not least, create a basis for the utility contract that can reduce some related balancing costs on the buyer's import utility contract.

This analysis shows that, despite the chaotic behaviour that makes wind forecast a very challenging task, a satisfactory forecast is achieved for one-year horizon. By breaking the problem down into several periods, with different parameters of beta distribution we could obtain a good forecast. A different approach was used for solar generation forecasting. The optimal combination of several probability densities provides a decision on which farms to choose for a particular consumption profile.

This paper is a simplified version of a real case scenario. In regular PPA negotiations, many projects would be offered several

generators with different price structures. The length of these contracts would lie between 10 an 25 years instead of one year, or shorter-term periods. However, this model still serves as a solid basis for real case scenarios, as they would use the same forecasting principles that can be extrapolated to a longer forecasting horizon and a higher number of assets.

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.

Data availability

The authors do not have permission to share data.

Acknowledgements

We would like to thank the Department for Business, Energy and Industrial Strategy of the United Kingdom, the Engineering and Physical Sciences Research Council DTP of Brunel University London, the National Physical Laboratory and Mitie for funding this research.

References

Abuella, M., Chowdhury, B., 2015. Solar power forecasting using artificial neural networks. In: 2015 North American Power Symposium. NAPS, IEEE, pp. 1–5.

Akçay, H., Filik, T., 2017. Short-term wind speed forecasting by spectral analysis from long-term observations with missing values. Appl. Energy 191, 653–662.

Azad, H., Mekhilef, S., Ganapathy, V., 2014. Long-term wind speed forecasting and general pattern recognition using neural networks. IEEE Trans. Sustain. Energy 5 (2), 546–553.

Bacher, P., Madsen, H., Nielsen, H., 2009. Online short-term solar power forecasting. Sol. Energy 83 (10), 1772–1783.

Barbounis, T., Theocharis, J., 2006. Locally recurrent neural networks for long-term wind speed and power prediction. Neurocomputing 69 (4–6), 466–496.

Bayes, T., 1763. A letter from the late reverend Mr.Thomas Bayes, FRS to John Canton, MA and FRS. Philos. Trans. R. Soc. Lond. 53, 269–271.

Billuroglu, B., Livina, V., 2022. Full-cycle failure analysis using conventional time series and machine learning techniques. J. Fail. Anal. Prev. 22, 1121–1134.

Bossanyi, E., 1985. Short-term wind prediction using Kalman filters. Wind Eng. 9 (1), 1–8.

Bracale, A., De Falco, P., 2015. An advanced bayesian method for short-term probabilistic forecasting of the generation of wind power. Energies 8 (9), 10293–10314.

Brindley, G., 2020. How can corporates minimise risks and maximise opportunities when entering renewable PPAs?: Re-source platform. http://resource-platform.eu/news/risk-mitigation-for-corporate-renewable-ppas/ (Accessed on 20 June 2022).

Chen, C., Duan, S., Cai, T., Liu, B., 2011. Online 24-h solar power forecasting based on weather type classification using artificial neural network. Sol. Energy 85 (11), 2856–2870.

Cheng, W., Liu, Y., Bourgeois, A., Wu, Y., Haupt, S., 2017. Short-term wind forecast of a data assimilation/weather forecasting system with wind turbine anemometer measurement assimilation. Renew. Energy 107, 340–351.

Constantinescu, E., Zavala, V., Rocklin, M., Lee, S., Anitescu, M., 2009. Unit Commitment with Wind Power Generation: Integrating Wind Forecast Uncertainty and Stochastic Programming. Technical report, Argonne National Lab.(ANL, Argonne, IL (United States.

Fulbright, N., Hedges, A., Currie, S., Baines, T., Nicols, A., OĎonovan, C., Dominy, P., 2021. Corporate renewable power purchase agreements: Scaling up globally. https://www.wbcsd.org/Programs/Climate-and-Energy/Climate/Resources (Accessed on 20 June 2022).

Gensler, A., Henze, J., Sick, B., Raabe, N., 2016. Deep learning for solar power forecasting—an approach using autoencoder and lstm neural networks. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics. SMC, pp. 002858–002865.

Gil, A., De La Torre, M., Domínguez, T., Rivas, R., 2010. Influence of wind energy forecast in deterministic and probabilistic sizing of reserves. In: 9th International Workshop on Large-Scale Integration of Wind Power Into Power Systems As Well As on Transmission Networks for Offshore Wind Power Plants.

Gilman, D., Fuglister, F., Mitchell, J., 1963. On the power spectrum of red noise. J. Atmos. Sci. 20 (2), 182–184.

Golestaneh, F., Pinson, P., Gooi, H., 2016. Very short-term nonparametric probabilistic forecasting of renewable energy generation—with application to solar energy. IEEE Trans. Power Syst. 31 (5), 3850–3863.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Hedges, A., Fulbright, N., Duvoort, M., 2019. How multi-technology PPAs could help companies reduce risk. https://www.wbcsd.org/Programs/Climate-and-Energy/Energy/REscale/Resources (Accessed on 20 June 2022).

Heinemann, D., Lorenz, E., Girodo, M., 2006. Forecasting of Solar Radiation. Solar Energy Resource Management for Electricity Generation from Local Level to Global Scale. Nova Science Publishers, New York, pp. 83–94.

Hölzle, U., 2017. Achieving our 100 renewable energy purchasing goal. https://www.google.com/green/pdf/achieving-100-renewable-energy-purchasing-goal.pdf (Accessed on 20 June 2022).

Huang, C., Liu, Y., Tzeng, W., Wang, P., 2011. Short term wind speed predictions by using the grey prediction model based forecast method. In: 2011 IEEE Green Technologies Conference. IEEE-Green, IEEE, pp. 1–5.

Jenkins, G., Lim, H., et al., 1999. An integrated analysis of a power purchase agreement. In: JDI Executive Programs. Technical report.

Jiang, Y., Song, Z., Kusiak, A., 2013. Very short-term wind speed forecasting with bayesian structural break model. Renew. Energy 50, 637–647.

Jones, E., Oliphant, T., Peterson, P., et al., 2001. SciPy: Open source scientific tools for Python. http://www.scipy.org/.

Juban, J., Siebert, N., Kariniotakis, G., 2007. Probabilistic short-term wind power forecasting for the optimal management of wind generation. In: 2007 IEEE Lausanne Power Tech. pp. 683–688.

Lange, B., Rohrig, K., Ernst, B., Schlögl, F., Cali, Ü., Jursa, R., Moradi, J., 2006. Wind power prediction in germany–recent advances and future challenges. In: European Wind Energy Conference, Vol. 1, no. 5. Athens, pp. 73–81.

Lowery, C., O'Malley, M., 2012. Impact of wind forecast error statistics upon unit commitment. IEEE Trans. Sustain. Energy 3 (4), 760–768.

Matos, M., Bessa, R., 2010. Setting the operating reserve using probabilistic wind power forecasts. IEEE Trans. Power Syst. 26 (2), 594–603.

Mauch, B., Apt, J., Carvalho, P., Small, M., 2013. An effective method for modeling wind power forecast uncertainty. Energy Syst. 4 (4), 393–417.

McLean, S., Gneiting, T., Raftery, A., 2013. Probabilistic wind vector forecasting using ensembles and bayesian model averaging. Mon. Weather Rev. 141 (6), 2107–2119.

More, A., 2022. Crypto (Bitcoin) mining by solar power — the next big thing! EQ Mag Pro, 10. https://www.eqmagpro.com/cryptobitcoin-mining-by-solar-power-the-next-big-thing-eq-mag-pro/ (accessed on 10 Oct 22).

Negnevitsky, M., Potter, C., 2006. Innovative short-term wind generation prediction techniques. In: 2006 IEEE Power Engineering Society General Meeting. p. 7.

Nikzad, A., Chahartaghi, M., Ahmadi, M., 2019. Technical, economic, and environmental modeling of solar water pump for irrigation of rice in Mazandaran province in Iran: A case study. J. Clean. Prod. 239, 118007.

Nikzad, A., Mehregan, M., 2022. Techno-economic, and environmental evaluations of a novel cogeneration system based on solar energy and cryptocurrency mining. Sol. Energy 232, 409–420.

Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Renné, D., Hoff, T., 2010. Validation of short and medium term operational solar radiation forecasts in the us. Sol. Energy 84 (12), 2161–2172.

Pinson, P., Chevallier, C., Kariniotakis, G., 2007. Trading wind generation from short-term probabilistic forecasts of wind power. IEEE Trans. Power Syst. 22 (3), 1148–1156.

Pinson, P., Madsen, H., Nielsen, N., Papaefthymiou, G., Klöckl, B., 2009. From probabilistic forecasts to statistical scenarios of short-term wind power production. Wind Energy: Int. J. Progr. Appl. Wind Power Convers. Technol. 12 (1), 51–62.

Reikard, G., 2009. Predicting solar radiation at high resolutions: A comparison of time series forecasts. Sol. Energy 83 (3), 342–349.

Salvatier, J., Wiecki, T., Fonnesbeck, C., 2016. Probabilistic programming in Python using pymc3. PeerJ Comput. Sci. 2, e55.

Sfetsos, A., Coonick, A., 2000. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. Sol. Energy 68 (2), 169–178.

Shi, J., Ding, Z., Lee, W., Yang, Y., Liu, Y., Zhang, M., 2013. Hybrid forecasting model for very-short term wind power forecasting based on grey relational analysis and wind speed distribution features. IEEE Trans. Smart Grid 5 (1), 521–526.

Solari, G., Repetto, M., Burlando, M., De Gaetano, P., Pizzo, M., Tizzi, M., Parodi, M., 2012. The wind forecast for safety management of port areas. J. Wind Eng. Ind. Aerodyn. 104, 266–277.

Stahley, B., 2019. Commercial solar panel degradation: What you should know and keep in mind. https://businessfeed.sunpower.com/articles/what-to-know-about-commercial-solar-panel-degradation.

Urquhart, B., Kurtz, B., Dahlin, E., Ghonima, M., Shields, J., Kleissl, J., 2015. Development of a sky imaging system for short-term solar power forecasting. Atmos. Meas. Tech. 8 (2), 875.

Walsh, B., 2004. Markov chain Monte Carlo and gibbs sampling. In: Lecture notes for Biostatistical Analysis, vol. 581.

Wilczak, J., Finley, C., Freedman, J., Cline, J., Bianco, L., Olson, J., Djalalova, I., Sheridan, L., Ahlstrom, M., Manobianco, J., et al., 2015. The wind forecast improvement project: A public–private partnership addressing wind energy forecast needs. Bull. Am. Meteorol. Soc. 96 (10), 1699–1718.