Highlights

A review of very short-term wind and solar power forecasting

Tawn, R.,Browell, J.

- Research in minutes- to hours-ahead wind and solar power forecasting is reviewed
- Knowledge transfer between wind and solar continues to benefit both fields
- An open-source case study of three distinct methods is presented
- Best practice in forecast evaluation is set out in the case study
- Common areas for improvement in research quality are identified

A review of very short-term wind and solar power forecasting

Tawn, R.[a,*], Browell, J.[a,b]

[a] Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK
[b] School of Mathematics and Statistics, University of Glasgow, UK

ARTICLE INFO

Keywords:
Forecasting
Wind Power
Solar Power
Renewable Energy
Probabilistic Forecasting
Open Data

ABSTRACT

Installed capacities of wind and solar power have grown rapidly over recent years, and the pool of literature on very short-term (minutes- to hours-ahead) wind and solar forecasting has grown in line with this. This paper reviews established and emerging approaches to provide an up-to-date view of the field. Knowledge transfer between wind and solar forecasting has benefited the field and is discussed, and new opportunities are identified, particularly regarding use of remote sensing technology. Forecasting methodologies and study design are compared and recommendations for high quality, reproducible results are presented. In particular, the choice of suitable benchmarks and use of sufficiently long datasets is highlighted. A case study of three distinct approaches to probabilistic wind power forecasting is presented using an open dataset. The case study provides an example of exemplary forecast evaluation, and open source code allows for its reproduction and use in future work.

1. Introduction

The increasing penetration of wind and solar energy in power systems around the world necessitate new ways of operating energy systems and markets. The variability and limited predictability of the wind and solar resource introduces uncertainty for planners and operators on all time scales, from seconds and minutes ahead, to decadal variability [1, 2] and climate change [3].

Forecasting plays a central role in minimising this uncertainty on operational time scales from real-time to a few days ahead [4]. Quantifying uncertainty is also necessary for 'optimal' decision-making and risk management. Forecast uncertainty is quantified in probabilistic forecasts which most commonly take the form of prediction intervals, predictive probability density functions (univariate or multivariate), or trajectories/scenarios, though other formats exist.

It is important to distinguish between *short-term* forecasting, with lead-times of hours to days ahead, and *very short-term* forecasting, with lead-times of minutes to hours ahead. The World Meteorological Organisation defines the *very short-term* range as up to 12 hours ahead [5], but in energy forecasting the distinction is generally methodological rather than at fixed lead time although neither convention is consistently applied. The term *nowcasting* is also used to refer to very short-term forecasting in the meteorology community, but here we will use *very short-term* throughout for consistency. The main source of predictability on short-term time scales comes from Numerical Weather Prediction (NWP), whereas the main sources of predictability on very short-term time scales are recent observations. NWP is not well suited to very short-term forecasting because of the time required for data assimilation and computation, and additional uncertainty introduced by weather-to-power conversion which is greater than natural variability on very short-term time scales. For the purposes of this review, which focuses on very short-term forecasting, we are concerned with forecasting methods based on recent observations and timescales where NWP adds limited or no value.

Wind and solar forecasts of the minutes and hours ahead are required by power system operators to manage the balance of supply and demand, and electricity market participants to trade energy. For instance in Denmark, the coun-

Acronyms: ANFIS, Adaptive Neuro-Fuzzy Inference System; AR, Autoregressive; AR(I)MA, Autoregressive (Integrated) Moving Average; CEEMDAN, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise; CRPS, Continuous Ranked Probability Score; ELM, Extreme Learning Machine; EMD, Empirical Mode Decomposition; IMF, Intrinsic Mode Function; LASSO, Least Absolute Shrinkage and Selection Operator; LSTM, Long Short-term Memory; MA(P)E, Mean Absolute (Percentage) Error; MC, Markov Chain; ML, Machine Learning; NN, Neural Network; NWP, Numerical Weather Prediction; RF, Random Forest; RMSE, Root Mean Square Error; SSA, Singular Spectrum Analysis; SVM, Support Vector Machine; SVR, Support Vector Regression; VAR, Vector Autoregressive; VMD, Variational Mode Decomposition.

*Corresponding author
rosemary.tawn@strath.ac.uk (T. R.); jethro.browell@glasgow.ac.uk (B. J.)
www.jethrobrowell.com (B. J.)
ORCID(s): 0000-0003-1298-9111 (T. R.); 0000-0002-5960-666X (B. J.)

try with the highest penetration of wind energy in the world, the Transmission System Operator's forecasts "are in five minute resolution and are updated every few minutes using all the latest information available" [6]. While countries such as Denmark are employing best practices such as leveraging real-time power production data, novel methods for producing increasingly accurate forecasts are continually being proposed, new sources of observational data are becoming available, and new ways of sharing data between parties are emerging. Current large collaborative forecasting projects include the European Smart4Res project [7] and several studies commissioned by the US Department of Energy's Solar Forecasting 2 program, such as the open-source solar forecast arbiter for forecast evaluation and benchmarking [8]. This article reviews these advances beyond current state-of-the-art operational forecasting systems and discusses their relative merits and potential evolution.

The expansion of wind and solar energy and research necessitates regular reviews and synthesis of advances, yet despite sharing many common features, wind and solar forecasting are often reviewed in isolation, perhaps a result of the relatively later development of solar power forecasting compared to wind [9]. Both wind speed and solar irradiance exhibit spatio-temporal correlation as a result of their dependence on large-scale meteorological phenomena. As such, some methods are effective for both wind and solar applications, such as time-series methods supplemented with exogenous inputs or multi-variate extensions which capture spatial correlations between multiple sites. In the recent history of very short-term wind and solar power forecasting one field has learned from the other. In this paper we identify potential opportunities for further advances in the same vein.

Both solar power forecasting [10, 11, 12, 13, 14, 15] and wind power forecasting [16, 17, 18, 19, 20] have been reviewed recently individually. However, very short-term horizons have received little attention in these reviews and neither have advances in very short-term wind and solar forecasting been compared. There are only two exceptions we are aware of: Sweeney [21] who consider wind, solar and hydro power together and discuss very short-term lead-times briefly, but do not systematically review the field and instead provide a vision for renewable energy forecasting in the future; and Barbieri [22] whose primary focus is very short-term solar but who also briefly mention transferable approaches from wind literature.

Previous reviews provide detailed analysis of various modelling approaches; for solar forecasting, Antonanzas [10] examines several approaches to persistence models for solar forecast benchmarks, Inman [12] covers clear sky models in depth, and Ahmed [14] includes particular detail on deep learning models and sky imaging. There have been several reviews of combined, or hybrid, models [23, 19, 16] while Giebel [17] provides an overview of the history of very short-term forecasting as well as models using NWP inputs. Jensen details a wide range of solar evaluation metrics, including for event-based forecasts, e.g. forecast performance for ramps [24]. Foley [18] gives average values for error metrics for different forecast horizons. Current state of the art and future directions suggested include greater prevalence of probabilistic forecasting [10, 25], increased focus on the economic impact of forecasts on decision making [10, 24], weather classification or regime-based approaches [14, 21], use of high resolution - including turbine level - data and data marketplaces [21]. Yang [15] uses a text mining approach to map forecasting and model terminology, before also highlighting six key recent works. Inman [12] identifies the forecasting of ramp events as a particular challenge for renewable energy integration in general. Key recommendations from these reviews include the need for a general database of geographically dispersed sites to test models on [26], consistent benchmarking approaches across research papers [11] as well as common evaluation metrics [10]. Lauret [27] recommends the Continuous Ranked Probability Score (CRPS) score for probabilistic forecast evaluation.

This review proceeds with a description of the systematic literature search that has been performed and a high-level bibliometric analysis (Section 2), after which very short-term solar (Section 3) and wind (Section 4) power forecasting are reviewed before a summary of common research methods and comparisons between the wind and solar literature are drawn in Section 5. While this review is by no means exhaustive, it is intended to give an overview of the variety of approaches that have been proposed in recent years. A case study based on an open dataset is presented in Section 6 in order to reproduce and compare three distinct classes of statistical model commonly employed for wind and solar forecasting but that are seldom compared to one another. The findings of this review and advances in very short-term wind and solar power forecasting are discussed in Section 7 which also speculates as to the direction of future research in this area.

2. Summary of papers reviewed

We used Web of Science to conduct a literature search[1], up to and including the end of 2020, for publications on short-term and very-short-term wind and solar forecasting. The number of works in this area has clearly been increasing substantially throughout the last decade, in line with the increases seen in both wind and solar generation globally (Fig 1). This suggests the importance of forecasting these variable generation technologies increases as their penetration on the grid increases [17].

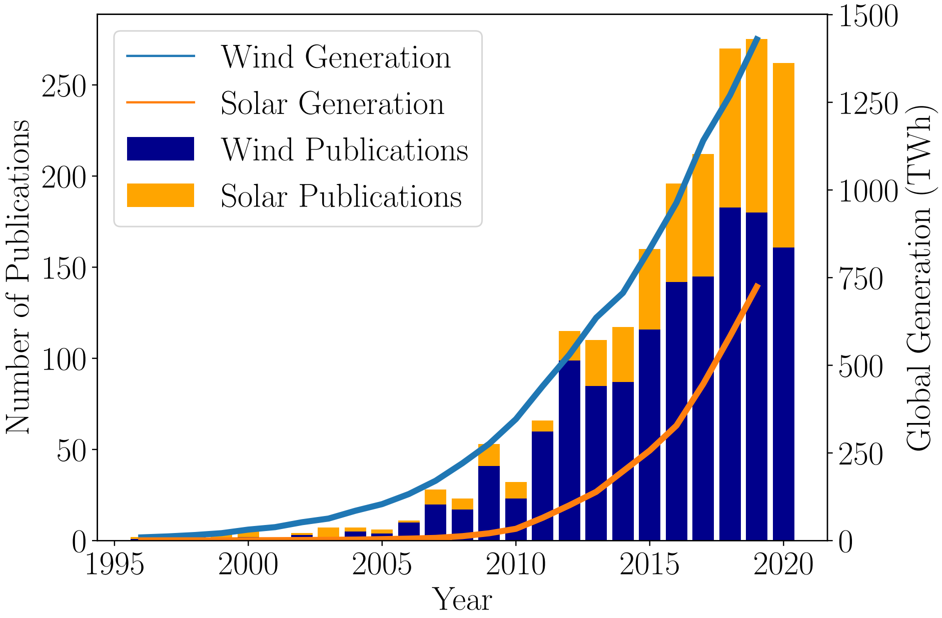


Figure 1: Forecasting publications broken down by wind and solar as a stacked bar chart, also plotted with global energy generation through time. Generation data provided under CC BY 4.0, Hannah Ritchie & Max Roser, https://ourworldindata.org/renewable-energy.

For short-term methods where lagged on-site measurements are the predominant data input, models often fall into two broad types: traditional time series regression, and Machine Learning approaches. Of the papers examined in this review, we found 24% included some type of regression or time series model, and 62% included a Machine Learning (ML) model. A list of all papers included is given in table A1 of the appendix. Figure 2 shows a general summary of forecasting approaches across the literature.

The subsequent sections 3 and 4 cover the top 50 most cited results stratified by the number of publications in each year and selected by the Web of Science search[1]. This selection has been limited to publications in 2014 or later, as the aim of this work is to focus on recent trends and developments in wind and solar forecasting. The literature from this search has also been supplemented with other references and works already known to the authors.

3. Solar power forecasting

Solar projects tend to have smaller installed capacities relative to wind projects: in the UK as of May 2020, the average solar installation has a capacity of 1.29 kW, with only 1.8% of these exceeding 4kW [28]. Of the larger UK projects requiring planning applications, the average installed capacity across 1171 projects was 7.2 MW, compared to an average of 29.6 MW across 778 wind projects [29]. As such, solar generation tends to consist of a greater number of smaller projects than wind. Sweeney et al [21] note that decentralised small scale energy sources often contribute to localised grid congestion problems, increasing the importance of accurate forecasts for grid management. Very small

[1] The search query used was ((TS=((("wind speed" OR "wind power" OR "solar" OR "renewable generation") NEAR/5 ("forecast*" OR "predict*")) AND ("short term" OR "short-term" OR "very-short-term") NOT("hydro" OR "thermal")))) AND LANGUAGE: (English)

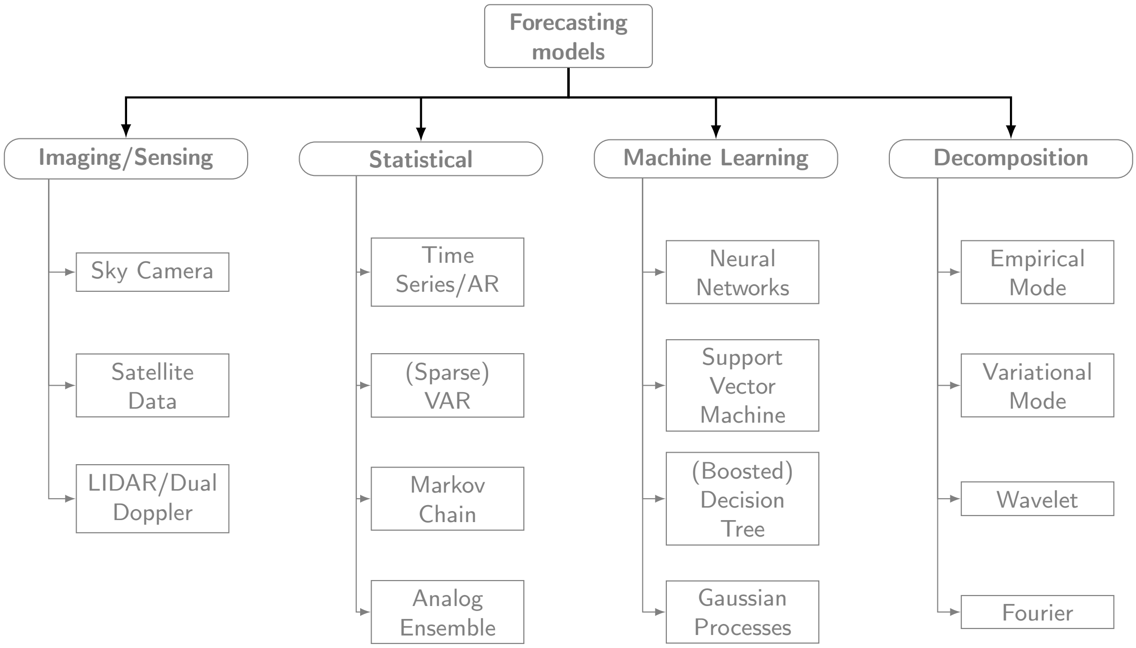


Figure 2: Diagram of forecasting model techniques. Neural Networks include ELM, RNN, CNN, LSTM etc; Decision tree methods include Random Forest and Gradient boosted trees. Methods may also be implemented in an adaptive or online framework, or include regime switching. They may also be used for probabilistic as well as deterministic forecasts.

solar systems such as household installations are often ‘behind-the-meter’, with no power production data available to forecasters and as such are often instead incorporated in ‘net demand’ (rather than power production) forecasts [10].

Solar power production follows strong seasonal and diurnal patterns due to the changing path of the sun, which defines the maximal possible irradiation for a given location, time and date. This is known as ‘clear sky’ irradiation, which can be well defined by various models [12]. In addition, the passage of clouds create shadows that introduce stochastic variability in the power time series and is much more challenging to predict [10]. Atmospheric aerosols may also reduce surface irradiation and therefore power output. This may be caused by natural phenomenon such as salt from sea spray, dust storms and soot from wildfires, or man-made pollution. A case study in West Africa found a reduction in power in the range 13-37% due to dust aerosols [30]. The physical condition of the panels can also affect production. For example, accumulation of dirt and dust have been shown to reduce energy production by 2-6%, and snow cover can also reduce power output completely if thick enough [31].

3.1. Image-based methods

Imaging techniques may be applied to either ground-based systems or satellite images to determine and predict future cloud cover, used in turn to forecast solar irradiance or solar power directly. Ground based sky imaging has mainly been used for high temporal resolution forecasts up to 30 minutes ahead. For a cloud at an altitude of 2km and a speed of 10 ms^{-1}, this represents a field of view of 154°. The focus of this method on very short time horizons is two fold: field of view and cloud formation and dissipation limit the skill of this method out to longer horizons [32], while it also fills a gap that several other data sources don’t currently have the spatial or temporal resolution to match [33] (satellite images generally have a 15 minute or slower update time for example). Methods using propagation of current observed cloud conditions are common such as cloud motion displacement [34] or determination and propagation of shadow position using cloud base height measurements in conjunction with images [35]. In this work clouds were also classified by type, although persistence still outperformed this method at a horizon of 25 minutes. Pitfalls of sky imaging systems may include errors due to perspective, image saturation in pixels close to the sun and soiling of the cameras [15]. There is also additional expense associated with maintaining a camera system on site.

Lago [36] trains an irradiance model using satellite and weather forecast data as inputs and ground measurements of solar irradiance at a group of sites in the Netherlands as the target variable. The learned model may then be used more generally at other sites without the need for ground measurements, and in fact this generalised model also outperformed models trained with local ground data. This approach is perhaps more suited to forecasting a group of sites rather than a single location, as a small subset of sites that do have ground-based measurements is also needed for model training. It would be interesting to test the generalisation of this approach to other climate regimes and more geographically dispersed sites. Harty [37] also uses both satellite and NWP data. However, they take a slightly different approach, producing cloud motion vector fields from both information sources and combing these via ensemble Kalman filter. Their method improves upon using a single information source for intra-hourly forecasts for a city region. Bellinguer [38] modelled spatio-temporal dependencies, with different models fitted conditional on NWP geopotential height. A combination of satellite data, where the 10 most informative pixels are chosen via mutual information, and on-site power measurements are used as inputs. Carriere [39] notes that different information sources tend to be most beneficial at different forecast lead times, so proposed an approach including several information sources. Irradiance time series from satellite data, NWP forecasts and lagged on-site power and temperature are supplied to the model, leading to good performance across a range of horizons up to 36 hours ahead. Non-parametric probabilistic forecasts were produced through an analog ensemble, using sets of similar past observations.

3.2. Probabilistic methods

Probabilistic methods allow quantification of uncertainty in the forecast and can facilitate proper risk analysis in applications. However, only a portion of the solar forecasting literature considers probabilistic forecasts and within this there is still sometimes a focus on general prediction intervals rather than full predictive densities.

Prediction interval approaches include a method using the variability of a time series about its mean [40]. Alternatively an ‘uncertainty metric’ may be determined from ensemble forecasts for points in a reference dataset, which is then used to look up the expected error (then used as a prediction interval) using a nearest neighbours approach [41].

Full density forecasts may be parametric, where the predictive distribution is specified by a small number of parameters (e.g. the mean and variance of a Gaussian distribution), or non-parametric with no assumed distributional shape. Golestaneh finds that solar forecast error distributions are not easily fitted by any common parametric distribution, so propose non-parametric quantile forecasts using lagged power alongside meteorological measurements in an improved Extreme Learning Machine (ELM) model [42]. This was demonstrated on a high resolution (1 minute) dataset which may not always be available. Gaussian process regression has also been proposed with an extension to give less weight to the observations that were more likely to be outliers [43].

3.3. Machine learning

Various machine learning techniques have been proposed for solar forecasting as they can allow for nonlinear relationships [44, 45] and learn from data without the need to make assumptions about the relationships between variables. For the very short term (up to one hour ahead), Rana [46] showed that on-site power measurements can provide skilful forecasts and NWP inputs (solar irradiance, temperature, humidity and wind speed) don’t further improve forecast skill. They used an ensemble of Neural Networks, which outperformed a Support Vector Regression (SVR) model. Feature engineering of a ‘cloud cover index’ from humidity and rainfall measurements and use of previous forecast errors as Neural Network (NN) inputs showed improved performance compared to a NN trained without these [47]. Long Short-term Memory (LSTM) networks are a common choice for time series problems; Lee [41] demonstrates their use with the dropout technique to produce ensemble forecasts. Alternative techniques to generate an uncertainty interval were also compared in this work. ELM models may overcome problems of overfitting and local minima associated with NN approaches. To reduce computational complexity, Majumder [44] used a low rank kernel ELM along with variational mode decomposition to address the nonstationarity of solar time series. This model was tested across a range of horizons (15 minutes to 1 day ahead). In other work using a cost function based on generalised correntropy for the ELM improved performance, possibly due to increased robustness to outliers [48]. Tang [49] also used an ELM to forecast solar power, in combination with pre-processing of inputs using an entropy method. The probabilistic approach of Golestaneh [42] is also based on ELM and performs favourably in comparison to both persistence and climatology as well as other ELM variants.

Abuella [50] used an ensemble of SVR models to generate day-ahead forecasts from NWP data; the 24 different forecasts are then combined to give the final probabilistic forecast via a Random Forest (RF). This approach shows improvement over individual models and could be appropriate for combining shorter-term forecasts. Eseye [51] also used NWP variables as inputs to an Support Vector Machine (SVM) model, additionally applying wavelet decomposition. However, the number of decomposed series were chosen based on previous literature rather than optimised on the given data.

For models with multiple data processing steps as well as model fitting, it may be advantageous to optimise all hyper-parameters for all parts of the model process simultaneously: [52] found a 53% improvement just by using simultaneous optimisation.

Spatio-temporal relationships are considered by including irradiance measurements from nearby sites as forecast inputs [45]. Not only is the proposed model shown to outperform Autoregressive (AR) methods, but boosted regression trees outperform both NN and SVR models. These models were developed only on times where clear sky irradiance exceeded a threshold, limiting their applicability to forecasts for dawn and dusk times.

3.4. Other methods

There has also been focus on utilising spatio-temporal dependencies between sites for solar forecasting. Agoua et al [53] propose a Vector Autoregressive (VAR) model normalising the input power time series by simulated power to make the time series stationary. They find Least Absolute Shrinkage and Selection Operator (LASSO) is the most effective variable selection procedure, and that conditioning on surface wind speed also adds skill to the forecasts.

The Sun4Cast system developed in the USA utilises several data sources and diverse models before producing a final forecast through a weighted combination [54]. The very short-term models include a sky imaging system, regression tree on pyranometer data, satellite imaging with advection and an NWP model tailored to solar forecasting with a high refresh rate [55]. They found benefits from each model for different lead times or climate scenarios, giving an effective combined model.

Several of the studies mentioned in previous sections may also be classified as hybrid methods. The term ‘hybrid forecast’ is often used to refer to methods where more than one forecasting method is combined into a final forecast: this may be simply through combining forecasts from different models [50, 41], applying some form of decomposition to the original time series and fitting different models to each resulting series [51], or where multiple different input data sources are processed separately before being combined [37, 55], for example satellite data and irradiation measurements. Hybrid methods often outperform a single model method, especially where a diverse set of individual models are combined. A full recent review of hybrid models for solar forecasting is given by Guermoui [23].

4. Wind power forecasting

The very nature of the wind presents forecasting challenges: the state of the atmosphere can never be fully known, meaning wind speed is treated as a stochastic process affected by many factors, from large scale weather systems down to local terrain. Of course the variable of interest in forecasting is often not power but wind speed. The relationship between wind speed and power is dynamic and nonlinear [56] which adds complexity and makes forecast power particularly sensitive to wind speed in between cut-in and rated wind speed. Wind power forecast errors are typically heteroscedastic and auto-correlated. Furthermore, production is bounded between zero and the rated capacity of a turbine or farm. These properties violate common assumptions in statistical modeling, such as independent and identically normally distributed errors, and should receive careful treatment in sophisticated forecasting methods.

Wake effects can influence the power output of turbines in the ‘shadow’ of others and this is highly related to wind direction. A power drop of around 30% of capacity was seen between the first and second row at Horns Rev when the wind direction is such that a turbine is directly behind another [57]. In cold climates, icing can reduce power output by as much as 40% [58]. Losses in operating efficiency over time could also affect forecast accuracy: turbine aging is estimated to cause a typical decrease in output of 0.2% per year in the first 5 years [59], although this also includes losses due to increased downtime. Data feed quality also affects the performance of models where site measurements are used as model inputs [60].

4.1. Regression-based methods

Past on-site measurements are widely collected by wind farm owners and are often valuable inputs when forecasting a few hours ahead. Simple time series methods based on Autoregressive moving average (ARMA) models are well established [17] and still form the basis of ongoing research. Zhou [61] shows that a dynamic combination of an Autoregressive integrated moving average (ARIMA) model with recent measurements as inputs and an AR model with inputs from NWP models is an improvement over either individual model. Other approaches using Autoregressive models in conjunction with other models are detailed in Section 4.4 on hybrid methods.

VAR models have been proposed to capitalise on spatial dependencies between geographically dispersed sites; since the number of model coefficients grows with the square of the number of sites, sparse models have been employed to reduce computational time and model complexity while improving forecast performance. For the case study presented by Cavalcante [62], a standard VAR model with no regularisation is shown to give improvement of around 5.9% over an AR model for a 2-hour ahead forecast, while introducing sparsity through LASSO regularisation gives a further 1% improvement. Grouping the LASSO penalty by whether an input is a lag of the predictor or not (i.e. diagonal vs off-diagonal elements) seemed to give the best results. An adaptive LASSO estimation algorithm is proposed in [63] to track potential changes in the VAR coefficients in an online fashion, yielding improvements relative to the equivalent static model for 15-minute resolution data and lead-times greater than 30-minutes.

Dowell developed probabilistic forecasts based on the logit-normal distribution in a VAR framework [64] for 5-minute ahead wind power forecasting; training on a window of most recent data allowed~~was used to allow~~ for changes in the sparsity through time. In a deterministic setting without the logit-normal transformation, and based on hourly mean powers, this method was outperformed by the LASSO-VAR approach. Correlation between farms has also been used to determine the sparsity of a VAR model [65], where the overall sparsity and number of non-zero coefficients for each farm can also be controlled. This was shown to outperform a standard LASSO-VAR model, but not compared against the sparsity structured LASSO in [62].

Capturing changes in VAR coefficients over time has been considered in adaptive frameworks where changes are tracked in an online setting [62, 64]. These adaptive methods improve over static equivalents, but inherently track changes with some lag and smoothing. Explicitly conditioning VAR coefficients on large-scale weather patterns was found to improve wind speed predictions from 1–6 hours ahead [66] but has not been applied to wind power.

For sites that wish to benefit from the improvements of spatio-temporal forecasting without revealing potentially commercially sensitive information, privacy preserving approaches have been developed. These may be grouped into three broad categories, each with their own disadvantages [67]: data transformation that may lead to a trade-off between privacy and model accuracy; multi-party computation [68] which may require a central coordinator and where similarity between model inputs and targets may lead to a breach in data confidentiality, or where using encryption techniques significantly increases computation time; and decomposition into parallel sub-problems which require iterative solutions - and each iteration progressively reveals more information to the participating data owners.

4.2. Machine learning

As with solar forecasting, various machine learning techniques have been applied to very short-term wind power forecasting. A comparison of SVR, decision trees and Random Forest models found Random Forest to give the lowest mean absolute percentage error [69] although no feature engineering was explored, which has been shown to play a significant role in good model performance in other works [70]. Correction of the output of an SVM using a Markov Chain showed improvement over a basic SVM approach [71]. The probabilistic output of a Markov Chain appears to be discarded in favour of a point forecast with 'fluctuation intervals'. A combination of two kernels (wavelet and polynomial) in an SVM model improved wind speed forecasts relative to the use of just a wavelet kernel [72]. The recent trend in wind speed (increasing,decreasing or stable) was also used to train separate models for these regimes, giving a slight improvement over a single model for all conditions.

A multi-objective approach was applied to NNs in [73], having separate objective functions for bias and variance. Similar multi-objective approaches have also been used on decomposed time series and are detailed in that section [74, 75]. Khodayar [76] used autoencoders for unsupervised feature learning and 'rough' neurons to better process noisy data, showing superior performance to other NN models. To its credit, forecast evaluation is based on a full year of out-of-sample data using the open source Western Wind dataset [77]. Neural Networks were also used ~~used~~ by Rodríguez [78] for 10-minute-ahead microgrid control.

Graph Neural Networks were used along with an LSTM for feature extraction to identify and utilise spatio-temporal relationships between sites by Khodayar [79], giving improvement over both persistence and other ML benchmarks. Inclusion of other metrics such as maximum observed error and correlation matrix of forecasts as well as usual average error metrics enhanced the analysis in this work, and the use of an open dataset is also a good step towards replication and comparison of research methods.

Hossain [80] also used convolutional NNs and Gated Recurrent Unit (GRU) layers for feature selection and processing of multiple input data sources respectively. They found improvement over other ML approaches at two case study sites.

De-noising of wind speed time series using Singular Spectrum Analysis (SSA) along with a fuzzy Neural Network model outperformed ARIMA and other NN implementations for a group of sites in China [81]. A novel neighbourhood LSTM network was proposed by Zhang [82] and claims to take causality, rather than just correlation, between variables into account, outperforming other ML methods in the study. Chen [83] compares artificial intelligence methods with Autoregressive models, finding that both an artifical Neural Network and an Adaptive Neuro-Fuzzy Inference System (ANFIS) marginally outperform an ARMA model for 10 minutes ahead forecasts, but that the ARMA model has superior performance for hour-ahead forecasts.

Many of the hybrid and decomposition approaches detailed in the following sections also make use of machine learning models.

4.3. Decomposition methods

Decomposition methods are based on the premise that the wind speed or power time series contain different frequency signals with different characteristics, and that modelling each of the decomposed series separately can lead to overall improvement in forecast skill [84, 85, 86]. Empirical Mode Decomposition (EMD) is based purely on the data and splits the original time series into several Intrinsic Mode Functions (IMFs), which can each have time varying frequency. As such, this method is applicable to nonlinear and non-stationary data [87].

Ensemble EMD, adding a noise term to the original signal before the decomposition, may be used to minimise mode mixing between the IMFs. Using ensemble EMD, Zhang [86] applied an ANFIS model to those IMFs classed as 'nonlinear' and a seasonal ARIMA model to those classed as 'periodic'. However, the judgement of which model to apply seems to have been made manually which may not be appropriate for real-world applications. Similarly, IMFs may be classed as high or low frequency signals, with different models applied to each; Liu [84], used an LSTM network for low frequency signals to capture longer-term trends, with an Elman NN for higher frequency IMFs. Similarly, a combination of ARIMA and NNs has been demonstrated to fit probabilistic forecasts to decomposed series [88]. An alternative approach fitted multiple different NNs to each IMF, with the final forecast for each IMF being a weighted combination of these [89].

To reduce the number of models estimated, Lu [85] used permutation entropy to group similar IMFs. An SVM was then used to forecast each series, outperforming both methods with no decomposition and those with decomposition but not using the permutation entropy approach. Decomposition has been combined with multi-objective optimisation for both accuracy and stability. This has been implemented with both Elman [74] and wavelet Neural Networks [75]. In both works the proposed methods outperformed single objective models.

Wavelet decomposition also results in the decomposition of a time series into multiple signals with different typical frequencies; it was found that further decomposing the highest frequency of these series improved forecasts [90]. Variational Mode Decomposition (VMD) is another decomposition technique, where each mode has a limited bandwidth. Zhang [91] found this outperforms EMD for the sites analysed.

4.4. Hybrid (combination) models

Hybrid models are based on the premise that a combination of several forecasts from different models, or where models use different information sets as inputs, commonly outperform a single model [92]. This does rest on the assumption that no model is the true representation of the underlying data generating process, as this single model, if known, would outperform any combination of 'misspecified' forecasts [93]. However, in many 'real-life' applications, either the true process is not known or no individual forecaster or model has access to the complete information needed to generate the 'perfect' model. This is certainly true of wind power forecasts, where the final value of power output is the result of complex physical interactions to produce the wind speed seen by the turbine, as well as the performance of the individual turbine and any imposed control actions.

The simplest method of forecast combination is a linear weighting approach where forecasts are combined as a simple weighted sum, often with the restriction of non-negative weights that sum to one. This approach was used for the combination of an SVM and radial basis function NN, where weights were found via forecast correlation with the actual time series for four different wind speed regimes for each month [94]. While specifying the model weights according to a correlation measure eliminates the need for estimation of the weights as free parameters, it may not guarantee the optimal combination.

Xiao [95] used linear weights to combine five different models, with the weights optimised both by minimising forecast errors ('traditional' approach) and using a particle swarm optimisation. Including all five individual models in the final combination consistently gave best results as opposed to dropping some model(s) completely, with the particle swarm optimised weights outperforming the traditional approach for this case. Zhou [61] found improvement using a small sliding window of previous forecast errors to adaptively combine forecasts, although only linear ARIMA type models were considered.

Nonlinear combination of an ensemble of neural network forecasts was achieved by a genetic programming algorithm [96]. Both lagged power measurements and NWP variables were used as inputs for one hour ahead forecasts, with feature selection to find the subset of 'informative' inputs although the results of this were not reported. Ouyang [97] takes a slightly different approach, determining significant input variables by Granger causality and building a separate univariate model for each of these. A multilayer perceptron was found to be best for combining the univariate predictions in the second stage of the model, and outperformed multivariate models. Lin [98] proposed a probabilistic forecast combination method, also using a weight coefficient for each model and combining both parametric and nonparametric forecast distributions. It is based on open data from GEFcom2014 [99]. Deterministic forecasts from a range of ML models have also been used as inputs for probabilistic combined forecasts [100]. While none of the individual models showed improvement over persistence for 1 hour ahead forecasts, the final combined model gave a significant (30%) improvement and beat persistence at all sites tested.

4.5. Probabilistic methods

A quantile loss function with an LSTM network was used to generate interval forecasts [101]. Attention mechanisms for automatic weighting of input features and extracting trends through time appear to improve the sharpness of the forecasts.

Jiang [102] used separate objective functions to maximise the interval coverage and minimise the interval width of a forecast power interval independently. This allows the user to choose from a set of pareto-optimal solutions according to their preferred trade-off between coverage and interval width. A deep learning approach using a convolutional Neural Network was found to outperform persistence and other shallow networks across seasons, quantile levels and for a wide range of forecast horizons [103].

A Markov Chain (MC) approach where transition probabilities between discrete power levels are modelled gives probabilistic forecasts without assuming a distributional shape [104]. A large number of power levels may lead to transition probabilities of zero in this method simply because they are not observed in the training data; a Bayesian approach where prior transition probabilities can be specified would mitigate this.

The Weibull distribution is commonly used to model wind speed distributions; Bracale [105] propose a mixture of two Weibull distributions to allow for bimodal distributions, fitting the mean with an ARIMA model and the remaining parameters through Bayesian inference. This approach outperformed both persistence and single distribution models for hour ahead forecasts.

The point forecast accuracy of an LSTM and the good probabilistic reliability of a Gaussian Process regression model were combined and found to outperform other time series methods both on point forecast accuracy and probabilistic performance [106].

4.6. Turbine-level data and remote sensing

Wind farms comprise multiple, sometimes hundreds, of individual wind turbines, forming a hierarchy which may be exploited to improve forecast performance [107]. Furthermore, if spread over a sufficiently large area, up-wind turbines may detect changes in wind speed early enough to inform very short-term forecasts for the farm as a whole. Similarly, measuring the wind speed up-wind of the wind farm using remote sensing may provide valuable information for very short-term forecasts.

Jiang [108] proposes use of time series from a neighbouring turbine and selection of forecast inputs via grey correlation analysis to improve individual turbine's forecasts. Along with an SVM model and cuckoo search for parameter optimisation, this does appear to improve forecasts relative to persistence, ARIMA and other SVM models. This model doesn't take account of the changing relationships between turbines as wind direction changes, for which a dynamic model may be more suitable.

A spatio-temporal Gaussian Process is proposed to predict turbine- and farm-level power production for 1- to 12-hours ahead [109], improving over non-spatial approaches to a comparable degree as spatio-temporal models on multiple wind farms. To its credit, this study is based on an open dataset [110]. One month of training data is used to train models on on a 6-hour rolling basis, which may impact some methods more than others.

Turbine-level forecasting using inputs from similar turbines (found through clustering algorithms) and an LSTM network showed promise over other ML benchmarks for 90-minutes ahead forecasts [111], although there was no discussion of how this translates to farm-level forecasts or consideration of hierarchical approaches for this.

Both lidar and radar technologies have been deployed at wind farms to measure the wind resource, though forecasting has not been the primary motivation. Wurth et al. [112] review minute-scale forecasting, with scanning lidar and radar identified as promising technologies; while use cases exist they are underdeveloped. Valledcabres [113] use dual doppler radar observations of up-wind wind speed to improve 5-minute ahead predictions of 1-minute mean power. Scanning lidar have also shown potential to improve forecasts for minutes-ahead horizons [114] but suffer from data reliability issues in fog or rainy conditions.

5. Research Methods

While very short-term power forecasting is an evolving area, certain methods are applied more commonly for different lead times. This is partly due to physical restrictions (for example cloud formation and dissipation and changes in wind direction limit the predictability of image based methods to long horizons) but also due to practical limitations, such as the latency in data assimilation, low temporal resolution, and low refresh rate typical of NWP models. However, higher spatial and temporal resolution NWP products are becoming available with hourly re-fresh rates, as provided by NOAA's High Resolution Rapid Refresh [115] and the UK Met Office's UKV [116] and MOGREPS systems [117], for example. Higher resolution and refresh rates are offered by emerging technologies such as Whiffle's so called 'finecasting' approach and NOAA's experimental 'Warn-on-Forecast' product [118]. These advances bring the ability of NWP to model and predict physical processes to ever shorter lead-times where they have not traditionally out-performed statistical methods based on local observations. The two approaches are complimentary, and state-of-the-art, site-specific forecasting systems combine both NWP and statistical processing of live site data.

Imaging techniques for minutes-ahead applications have had greater attention in the solar literature, whilst models including spatio-temporal relationships have focused more on wind power forecasting. Methods producing a forecast as a probability distribution are becoming more widespread, although there is more focus on probabilistic forecasting in the wind community. Solar forecasts sometimes only give a single confidence interval which might not have a formal definition in terms of probability coverage [40, 41]. Probabilistic forecasts are not always evaluated using probabilistic metrics, or only for one interval rather than the whole distribution [43, 47, 101].

Confidence in the significance of results may be undermined by use of limited case study datasets with a length of days to weeks rather than a year or more [119]. In particular, results of model evaluation carried out entirely on data from one season at one site may not generalise to other seasons, weather conditions or other locations. The shorter the dataset, the smaller the probability the data contains a wide range of weather (cloud or wind) conditions; this increases the risk of poor performance when forecasting for conditions not included in the training set. A long dataset covering multiple sites would be expected to allow more robust conclusions on model performance to be drawn. Use of small datasets is seen in both solar [47, 49, 52] and wind [71, 72, 81, 86, 84, 95, 106, 108, 88] studies.

Papers on novel methods do not always include appropriate benchmarks such as naive models or established best-in-class methods; we found both solar and wind papers which only compare models to their own variations [40, 41, 44, 49, 61, 71, 72, 73, 78, 102, 104]: this is in line with a survey by Doubleday et al [11], who find that 8 of 42 solar forecasting papers surveyed did not include a benchmark other than variants of the same model. They recommend comparison to two benchmarks, one highly reliable but more naive approach and one closer to state-of-the-art. Testing against benchmarks that are significantly different from the proposed model would allow for a clear comparison with other methods. Consistent benchmarks across papers and publishing code alongside for reproducibility would not only strengthen confidence in reported results, but allow easy comparison of state-of-the-art approaches.

We have found it is sometimes unclear how, or if, data has been partitioned to perform out-of-sample evaluation [69, 82, 89, 74, 75, 90, 94]. A brief clear description of the training and testing sets or cross validation approach used would be beneficial in these cases.

For wind power forecasting, a proportion of work is based on wind speed, rather than power, forecasts [66, 76, 88]: while wind speed forecasts may well be more appropriate for some applications, it is worth noting that grid or trading decisions require power forecasts. The conversion from wind speed to wind power is complex and nonlinear in itself and so models reporting skill in forecasting wind speeds are not guaranteed to provide the same level of skill if used to forecast power instead. Likewise, papers based on wind speed datasets where a power variable has been simulated by passing values through a power curve will not be representative of the noisy power data seen operationally [78, 83]. Open source wind power datasets [99, 9] now allow testing of models on power (rather than wind speed) data when that best fits with the aim of the study. For solar power, solar irradiation forecasts are analogous to wind speed forecasts in that they forecast a proxy for power, but not power itself. Although the relationship between irradiance and power output for solar is less complex than the wind speed-power relationship, results of studies based on irradiance forecasts are not guaranteed to generalise to power forecasts.

One of the strengths of Machine Learning based research papers seems to be the general prevalence of data pre-processing including data cleaning and variable transformations such as principal component analysis [79, 81]. Pre-processing is also seen in ML approaches to solar forecasting [47, 49].

6. Case study

The following case study is provided, along with underlying data and code, to serve as an example of good practice, and to highlight features of different approaches to very short-term forecasting. Three distinct methods found in the preceding literature review are implemented and evaluated using data from the wind track of GEFcom2014 [99].

The Vector Autoregressive (VAR) method was chosen to demonstrate modelling of spatio-temporal dependencies for a set of sites, while the Markov Chain (MC) method is a computationally simple nonparametric approach where no distributional shape is assumed. Finally, the decomposition (Empirical Mode Decomposition (EMD)) approach was chosen as a contrasting method and is often only benchmarked against other decomposition approaches. The GEFcom2014 data is publicly available and comprises two years (2012 and 2013) of hourly resolution data, including both wind power measurements and Numerical Weather Prediction (NWP) forecasts of wind speed at 10 and 100m for ten wind farms in Australia. However, here we only use the wind power and lagged values thereof for very short-term methods. The data is separated into the same training and testing sets for each method to allow fair comparison of the forecast errors. The first 1/3 of the data (up to 2012-08-31) are used for model training and hyperparameter optimisation while the remaining 16 months are used for forecast evaluation. Forecasts are generated for 1 to 6 hours ahead.

All data and code for this case study is available at doi.org/10.5281/zenodo.5070758 for reference or to use as benchmarks for other research. All methods implemented in this case study produce probabilistic forecasts to allow full description of the forecast distributions. For the Vector Autoregressive (VAR) and Empirical Mode Decomposition (EMD) approaches, parametric probabilistic forecasts were produced by log-transforming the data and fitting a Gaussian distribution to the transformed values. The log transform is defined as

$$y = \ln\left(\frac{x}{1-x}\right) \quad , \quad 0 < x < 1. \tag{1}$$

Power values x in the range $[0, \epsilon]$ are rounded to $x = \epsilon$ (and likewise for high powers $[1-\epsilon, 1]$ rounded to $x = 1-\epsilon$) to prevent infinities in the transformed data. This is effectively the same as putting all the probability mass from these boundary regions at ϵ and $1-\epsilon$. The mean of the distribution of the transformed values is modelled with VAR or EMD, and the variance of the distribution is also modelled to fully specify the Gaussian predictive distribution (in transformed space). Both a constant value for the variance, and a simple exponential smoothing model, were tested and the approach that minimised the pinball score chosen at each forecast horizon. Once the mean and variance are specified, quantile forecasts may be generated from the forecast Gaussian distribution, before applying the inverse transform

$$x = (1 + e^{-y})^{-1} \tag{2}$$

to produce quantile forecasts for the wind power.

6.1. Vector Auto-Regression

A simple vector autoregressive model with Least Absolute Shrinkage and Selection Operator (LASSO) has been implemented, using the log normal transformation in the same way as in [64]; after transformation, the data is modelled as normally distributed and forecasts are defined by the mean and variance of this distribution. In the transformed space, the mean of the distribution is modelled by a vector autoregressive process whereby the previous m lags from all sites are included as inputs and the regularisation parameter λ allows control over the strength of regularisation (and therefore the number of non-zero input coefficients) in the model. λ and m were optimised through cross validation on the training set, before the final model was fitted on the training data and used to forecast for all of the test set. The value of ϵ was set to 0.01. The variance of the forecast distribution was modelled as constant, and found using the variance of the residuals for the training folds (found via cross validation). An exponential smoothing model for the variance was also tested but did not provide any improvement over the constant variance model.

6.2. Empirical Mode Decomposition

Empirical Mode Decomposition allows a signal to be separated into different sub-series (called Intrinsic Mode Functions (IMFs)), each with different characteristic frequencies. Because it is an empirical approach based on the data, it allows for time-varying frequencies in the decomposed series and, by definition, the individual IMFs (plus the final residual) sum to the original signal for all time points. This allows different models to be fit on the different series, with the aim of improving the overall model fit. For example, one model may be better suited to forecasting the higher frequency IMFs, and a different model for the low frequency ones.

Following [86] we chose Autoregressive (AR) and Adaptive Neuro-Fuzzy Inference System (ANFIS) models as candidate models for each IMF. We fit a separate model to each IMF with no grouping. We implement Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to minimise mode mixing and to improve the spectral separation of modes [120].

A sliding window approach was implemented, with sliding windows over the training data used to find the optimal number of lags to include as inputs in the ANFIS model, the best performing model for each IMF (AR or ANFIS), the optimal number of IMFs and the optimal window length. Finding the 'best' model for each IMF relies on the assumption that better IMF forecasts will also sum up to a better overall forecast for the original series. CEEMDAN was applied to each window separately as the values of individual IMFs can change, particularly at the boundaries, for different windows. Although this is computationally more expensive than decomposing the entire series at once and all decomposed series are still guaranteed to sum to the original time series under this simpler approach, decomposing the series separately for each window guarantees no 'information leakage' from future values occurs in the decomposition. Once the types of model (AR or ANFIS), number of IMFs to use and sliding window length has been found, forecasts can be generated for all sliding windows in the testing set.

We found that the ANFIS model performed worse than an AR model for all IMFs. It is possible that increasing the number of inputs to the ANFIS model would give improvement, but this also results in a significantly higher computational cost. Five or six IMFs were found to give the best results.

It should be noted that due to the computational time associated with fitting several different models to all IMFs for many windows, choice of the optimal model for each IMF and optimisation of parameters was only done for one step ahead forecasts, rather than optimising separately for every forecast horizon.

6.3. Markov Chain method

Similar to AR methods where the forecast is dependent on previous values, Markov Chains assume the Markov property: the state at time $t+1$ depends only on the state at time t. Power values are discretised into a finite number of states and the transition probabilities between states result in a forecast probability for each power state, given the previous power observation. These probabilities may then be converted into quantiles to produce a nonparametric probabilistic forecast distribution. A frequentist approach to building a MC model would involve the calculation of a 'transition matrix' of probabilities of transitioning from each (discrete) state at time t, to each state at time $t+1$. The maximum likelihood estimates for transition matrix entries are found using counts of transitions between states from the training data [104]. This produces a transition matrix entirely dependent on the observed training data and may lead to transition probabilities equal to zero between certain states, simply because that transition was not observed over the training period. The uncertainty on the transition matrix entries is not accounted for. A Bayesian approach introduces priors to help account for this; the end forecast probabilities for each state are then effectively an integral over all possible values for the transition matrix entries, taking prior estimates and observed transitions into account. We follow [121] and use a Dirichlet prior which allows for a neat analytic solution. In the Bayesian formulation, the forecast distribution is

$$p(y|x) \propto \int p(y|\Theta)p(\Theta)p(x|\Theta)d\Theta \tag{3}$$

where x is the training data and Θ are the transition matrix values: θ_{ij} represents the probability of transitioning from state i to state j. $p(\Theta)$ are the prior probabilities, given by the Dirichlet distribution; for a MC with K discrete states the l^{th} row is given by

$$p(\Theta) = \frac{1}{B(\alpha)} \prod_{j=1}^{K} \theta_{lj}^{\alpha_j - 1}. \tag{4}$$

$p(x|\Theta)$ is the likelihood function:

$$p(x|\Theta) = \prod_{i=1}^{K}\prod_{j=1}^{K} \theta_{ij}^{n_{ij}} \tag{5}$$

The value of α_{lj} has to be specified for each lj element in the transition matrix, i.e. there are K^2 prior values for a MC with K discrete states. It is reasonable to assume that a transition to a more similar (closer) state is more likely than a large jump in power between time steps, so we constrained the prior values to adhere to this by defining

$$\alpha_{lj} = K - |l - j|. \tag{6}$$

To be able to optimise the importance of the observed data relative to the priors, a 'scaling factor' c was also introduced. For a forecast input state l, the final forecast probabilities are then

$$p(y|x) \propto \begin{pmatrix} cN_{l1} + \alpha_{l1} - 1 \\ cN_{l2} + \alpha_{l2} - 1 \\ cN_{l3} + \alpha_{l4} - 1 \\ \ldots \\ cN_{lK} + \alpha_{lK} - 1 \end{pmatrix} \tag{7}$$

Finally, this vector is normalised so that the elements sum to 1 (i.e. the total probability across all states is one).

Transition probabilities may change over time, so a sliding window using only the most recent data points was employed. The length of this sliding window was optimised over the training set, as well as the number of discrete states K and the scaling factor c giving the relative importance placed on the counts versus the priors. Forecasts could then be produced for the testing period.

6.4. Persistence

Persistence models are a common benchmark for time series forecasting methods [17], as they are very simple but often hard to beat on short time scales. Forecasts based on a Gaussian distribution are used, where the mean is equal to the most recent observation and the standard deviation is found from the standard deviation of residuals in the training set:

$$\hat{y}_{t+k|t} \sim \mathcal{N}(y_t, \sigma_t) \quad , \quad \text{where} \quad \sigma_t = \sqrt{\frac{1}{T}\sum_{t=k+1}^{T}\left(y_t - y_{t-k}\right)^2} \quad . \tag{8}$$

6.5. Forecast evaluation

When developing new forecasting methods and tools it is necessary to establish some criteria by which success and improvement upon existing practice are defined. Error metrics such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) are regularly used in academic literature and practice, with the forecast with the best score declared *the best*. In this case study, as well as exploring three distinct forecasting methods from the literature, we also offer an exemplary comparative evaluation of their performance. We focus on quantitative evaluation, following [119] in particular, but also comment on other important qualitative issues, such as computational time and interpretability [122]. In what follows we briefly introduce the metrics and scores we will employ, and direct readers to [119, 122, 24] and other sources referenced therein for more detailed discussion.

6.5.1. Evaluating deterministic forecasts

We employ two established metrics to evaluate deterministic forecast performance: MAE and RMSE. In both cases, metrics are defined for specific lead-times k steps ahead. For T forecasts $\hat{y}_{t+k|t}$, $t = 1, ..., T$ of y_t made k steps ahead at time t, MAE and RMSE are given by

$$\text{MAE}_k = \frac{1}{T}\sum_{t=1}^{T} |y_{t+k} - \hat{y}_{t+k|t}| \quad , \quad \text{and} \tag{9}$$

$$\mathrm{RMSE}_k = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_{t+k} - \hat{y}_{t+k|t})^2} \quad . \tag{10}$$

MAE is favoured by many practitioners due to its ease of interpretation, whereas those in the modelling community may prefer RMSE as it is analogous to the loss function used in model estimation when the objective is to produce conditional expectations as forecasts. Similarly, MAE corresponds to the conditional median. In any case, the ranking of forecasts by performance would not change if one metric were used instead of the other.

6.5.2. Evaluating probabilistic forecasts

Probabilistic forecasts are evaluated according to the principle of minimising sharpness subject to reliability. Reliability is verified using reliability diagrams, and then reliable forecasts may be discriminated using the Pinball Loss (also known as the 'quantile score'). Pinball loss for probability level α and lead-time k is given by

$$\mathrm{Pinball}_{\alpha,k} = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{q}^{(\alpha)}_{t+k|t} - y_{t+k}\right)\left(\mathbb{1}(y_{t+k} \leq \hat{q}^{(\alpha)}_{t+k|t}) - \alpha\right) \tag{11}$$

where $\hat{q}^{(\alpha)}_{t+k|t}$ is the predictive quantile of y_{t+k} with probability level α made at time t. The Pinball Loss is typically averaged across probability levels α to give a single score for the forecasting method being evaluated.

6.5.3. Comparing forecasts

Skill scores are a useful way of comparing forecasting methods because they are unit-free. For a given metric, the *skill* of the candidate method with performance metric M relative to a reference method with performance M_{ref} is given by

$$\frac{M_{\mathrm{ref}} - M}{M_{\mathrm{ref}} - M_{\mathrm{pref}}} \tag{12}$$

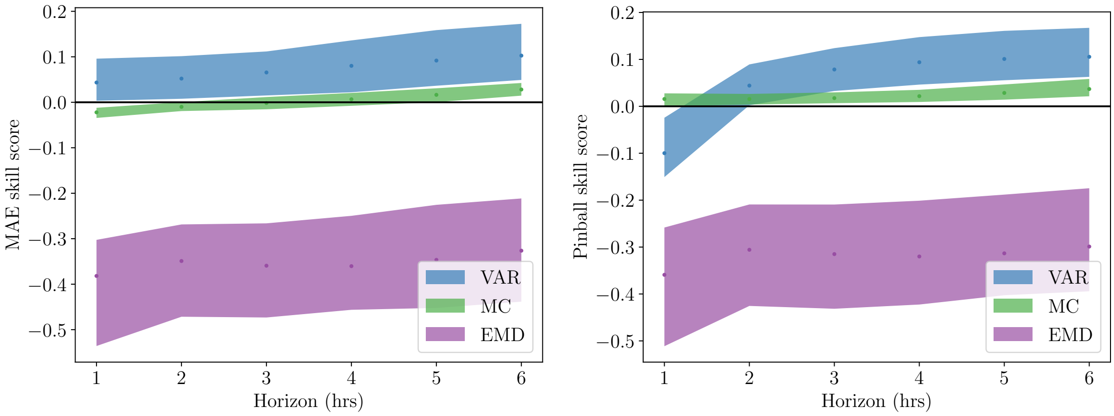
where M_{pref} is the 'perfect' score for the given metric, which is zero in many cases. A skill score of 0 indicates no improvement relative to the reference method and positive skill score indicates superior performance. A common reference method in wind and solar power forecasting is persistence (or *smart* persistence in the case of solar) and provides a robust benchmark for very short-term forecasts. On longer lead-times climatology is a more common reference method, e.g. the seasonal average for a given time of year.

A practical limitation on forecast evaluation is the finite number of forecast-observation pairs available when calculating metrics and skill scores. As a result, it can be difficult to establish whether any observed difference in performance will generalise or whether it is the result of sampling variation. Bootstrapping is a popular non-parametric method for quantifying the impact of sampling variation [123] involving re-sampling forecast errors and calculating error metrics multiple times in order to estimate sampling variation. If variation in metrics/skill scores overlap then any difference in performance is unlikely to generalise and may be a result of sample variation. Additionally, bootstrapped skill scores provide greater discrimination than independently re-sampled metrics [119]. However, care must be taken where serial correlation is present, which is often the case in forecasting tasks. Failure to account for serial correlation, e.g. by employing a block bootstrap, would result in over-confident results. Alternative tests for the significance are also available, such as the Diebold-Mariano test [124].

6.5.4. Results

Skill scores were calculated for all zones, using timestamps with forecast-observation pairs available for all zones and averaging the skill score from each zone. Figure 3 shows the variation in skill score with forecast horizon for all three models in this case study, relative to probabilistic persistence. For deterministic measures (MAE and RMSE) – which are based only on the q50 value – the VAR model outperforms both persistence and the other models tested, for all horizons. However, the MC model has the best pinball score for one step ahead forecasts, perhaps due to its nonparametric nature and therefore lack of distributional assumptions. The decomposition (EMD) approach is significantly worse than persistence for all horizons.

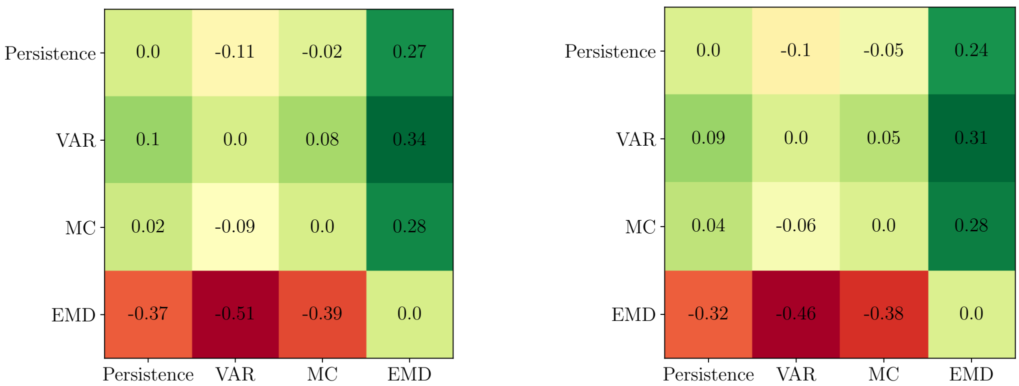
It is also beneficial to compare each model to each other model; matrices of the mean skill score between models are presented in Figure 4 for a 2 hour ahead horizon. A positive value indicates the model on the y-axis outperforms



(a) MAE skill score against forecast horizon

(b) Pinball skill score against forecast horizon

Figure 3: Skill scores of case study models, relative to probabilistic persistence model. The 95% interval of bootstrap samples is shown. Positive values indicate improvement over persistence. VAR=Vector Autoregressive, MC=Markov Chain, EMD=Empirical Mode Decomposition.



(a) Matrix of skill scores of MAE

(b) Matrix of Pinball skill scores

Figure 4: Matrices of MAE (left) and Pinball (right) skill scores for the 2h-ahead forecast produced by all combinations of models implemented. A positive value indicates the model on the y-axis outperforms that of the model on the x-axis. The VAR model outperforms all others in terms of both MAE and Pinball metrics at this horizon. VAR=Vector Autoregressive, MC=Markov Chain, EMD=Empirical Mode Decomposition.

that of the model on the x-axis. This clearly shows the EMD approach has the most extreme skill scores, whereas the relative performance of the other models are closer. The VAR model is the only one to outperform all other models. RMSE shows almost identical skill scores to MAE, and pinball skill scores are also similar.

For probabilistic forecasts, the best forecast should be sharp, subject to reliability. This cannot be judged from a single score value such as pinball loss, and so reliability diagrams also play an important role in probabilistic forecast evaluation. Relative Empirical frequency has been plotted, so that a perfect forecast would have a value of zero. For example, it would be expected that in a perfect forecast distribution, the observed power would be less than the q20 quantile forecast 20% of the time and the difference between expected and observed frequencies (the relative empirical value) would be zero. Figure 5 shows the reliability across the q5-q95 quantiles for the case study models. Both persistence and to a lesser extent the VAR forecasts display the s-shaped curve associated with too broad a forecast distribution, while the MC and EMD forecasts show bias (under and over-forecasting respectively). The confidence intervals derived from bootstrap resampling show the deviations from 'perfect' reliability are significant for all models.

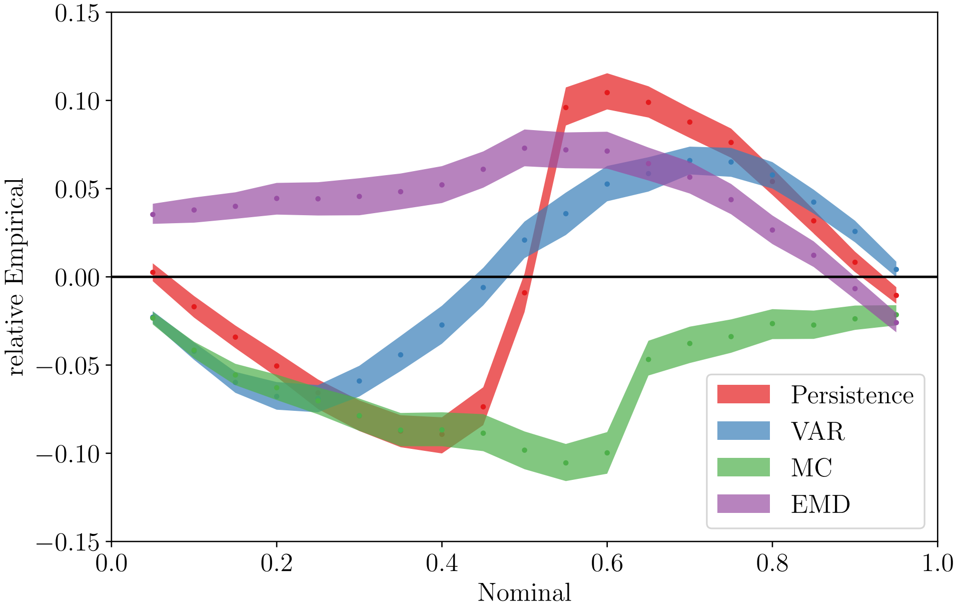


Figure 5: Relative reliability of two hour ahead forecasts for the case study models at zone 4. A relative empirical frequency of zero represents ideal reliability. VAR=Vector Autoregressive, MC=Markov Chain, EMD=Empirical Mode Decomposition.

6.6. Summary

This case study shows examples of contrasting methods for very short-term wind power forecasting and their relative performances, with the VAR approach proving the most skilful. This is perhaps not surprising given it is the only model to use inter-site dependencies in the forecasts. The Markov Chain model produces a nonparametric forecast, meaning no prior knowledge or assumption of the forecast distribution is needed. It shows superior performance for one step (one hour) ahead forecasts, but its skill is lesser for longer horizons, likely due to the fact it only uses one forecast input (the lag one power value). Decomposition models seem unlikely to provide competitive performance to other methods unless very different models are optimal for the different IMFs, and it can be unclear how best to choose which model to fit to each series. Grouping the IMFs before model fitting was not explored in this work; while this may improve forecast performance, it requires an additional step of forecast setup tuning which would significantly increase the time and effort needed to optimise the forecast setup.

The MC model is the most computationally fast of the models, effectively only requiring to discretise the power values and count the number of transitions between each level. In this study we have fixed the structure of the priors and only tuned their strength relative to the observations, as tuning each prior individually would be much more complex, but other structures could also be explored. While the VAR model takes slightly longer to fit (around a second to train once, tested with 25 different regularisation strengths for 6 different numbers of lags, for each forecast horizon), it fits one model for all locations simultaneously and predicting from a fitted model is still fast. However, the EMD model is significantly more complex both to train and to predict from, due to the additional decomposition step and then models fit separately for all of the decomposed series.

None of these models are perfect, but are intended to serve as open source examples for benchmarking future research and a demonstration of good practice in forecast evaluation. While this case study is only demonstrated for wind data, the code has been made available and could also be applied to solar data, although care must be taken when preparing the data to account for the diurnal and annual cycles~~diurnal and annual seasonality~~. Models that augment inputs (e.g. sky images and other weather data) show improvements [21] but such data were~~was~~ not available here.

7. Discussion and future work

Demand for ever more accurate very short-term wind and solar power forecasts has motivated a growing volume of research over the past decade, a trend which shows no signs of slowing. The vast majority of published research focused on wind power in the first half of the decade, but solar has been catching up and in 2019 there was one solar publication for every two in wind. In both cases there has been a shift to probabilistic forecasting, with authors citing benefits for users that become more acute as penetration of wind and solar increases.

The parallel development of very short-term wind and solar forecasting has benefited both fields. Approaches initially developed for wind power, such as exploiting spatio-temporal dependency, have been successfully adapted for solar. Similarly, the relatively well established use of remote sensing data in solar is showing potential for wind. Satellite images capture cloud motions on relevant time scales offering significant benefit for solar forecasting, but no equivalent for surface wind speeds has been demonstrated. A comparison may also be drawn between sky cameras and LIDAR; both are dedicated hardware for measuring the approaching solar or wind resource, respectively. Sky cameras are established tools for very short-term solar forecasting whereas only a few examples of scanning LIDAR for wind exist, likely due in part to significant differences in hardware and maintenance costs. If a sufficient economic incentive (or regulatory necessity) emerges for more accurate very short-term wind power forecasts, remote sensing may represent a suitable opportunity for forecast improvement.

In addition to theoretical advances, practical considerations have been the subject of recent research, including handling quality issues and data sharing. Data quality may be compromised by communications failures or operator actions, such as curtailment or integration with co-located storage. When a wind or solar farm is curtailed or metered alongside a co-located battery, its power output is no longer representative of local weather conditions with negative consequences for training forecast models and operational forecasts based on live power data. Where there is a broad literature on this topic in general, application to very short-term wind forecasting has only been considered in [60]. A related challenge which has received almost no attention in the literature is the prediction and/or utilisation of *power available signals* from curtailed wind and solar farms on very short-term forecast horizons. Plant controllers can typically produce accurate predictions of present power available but not forecasts of future values. Data sharing between wind and solar farms is necessary in order to capitalise on spatio-temporal information for very short-term and regional forecasting. Some data owners prefer not to share data they consider to be commercially sensitive or private, but it may be possible for them to do so in such a way that improves forecast performance while preserving privacy. In the absence of open data or a central forecast provider, privacy preserving sharing for spatio-temporal very short-term wind power forecasting was first proposed in [125], later developed in [68], and recently reviewed in [67]. Furthermore, data markets have been proposed to provide a financial incentive to share data in this way, introduced in the context of renewables in [126]. Further development is required to refine such algorithms, which can be demanding in terms of both computation and communication requirements, to develop compelling business models for data markets, and to ensure that they are cyber-secure.

Time-series based methods for both wind and solar forecasting have benefited from contributions from a range of disciplines including statistics, signal processing and machine learning among others. The application and adaptation of optimisation techniques capable of scaling to high-dimensional time series prediction is a good example of this. The significance of proposed methods risks being undermined when case studies are evaluated on small private datasets (hours to days, rather than months to years) and only compared to variations on the same approach. Often methods are evaluated for wind speed forecasting and their suitability for application wind power forecasting is not discussed or verified. Guidance and recommendations for forecast evaluation, including dataset size and properties, benchmarks and significance testing may be found in [119, 11, 122].

The reproducibility of energy forecasting research has improved in general over the past decade with use of open datasets and publication of code becoming more common. The Western Wind dataset [77] covers a large number of US locations, but power data is simulated (using wind speeds and a manufacturer power curve) rather than direct measurement; for this reason it might not be the most appropriate dataset to validate forecasting models on. A number of forecasting competitions have also released datasets, notably the GEFcom series which also publish descriptions of top performing methods and their performance, which may serve as benchmarks for future innovations. However, none of these competitions have featured very short-term forecasting to date, instead focusing on day-ahead time scales where the main task is post-processing numerical weather predictions. Competition formats based on providing training data comprising input-output pairs and a test set of only inputs (with corresponding output held by the organisers for evaluation) does not translate well to time series forecasts where lagged values are a necessary input. A competition focused on very short-term forecasting would be more challenging to run (e.g. running truly live, or requiring participants to submit software) but could make a valuable contribution to the field.

Acknowledgements

Rosemary Tawn is supported by The Data Lab Innovation Centre funding with co-funding from Natural Power Consultants Ltd. Thanks to Robert Graham and anonymous peer reviewers for helpful comments on drafts of this paper. Thanks also to the organisers of GEFcom2014 for the release of competition data.

Data availability

The data and code for the case study can be accessed at doi.org/10.5281/zenodo.5070758. The data here is a modified version of that provided for GEFcom2014 [99].

References

[1] J. Wohland, D. Brayshaw, H. Bloomfield, M. Wild, European multidecadal solar variability badly captured in all centennial reanalyses except CERA20C, Environ. Res. Lett. 15 (2020) 104021.

[2] J. Wohland, N. E. Omrani, N. Keenlyside, D. Witthaut, Significant multidecadal variability in German wind energy generation, Wind Energy Sci. 4 (2019) 515–526.

[3] D. Hdidouan, I. Staffell, The impact of climate change on the levelised cost of wind energy, Renew. Energy 101 (2017) 575–592.

[4] George Kariniotakis, Renewable Energy Forecasting: From Models to Applications, Woodhead publishing, 2017.

[5] W. M. Organisation, https://www.wmo.int/pages/prog/www/DPS/gdps.html, Accessed 09/11/2020.

[6] M. Baldini, The Danish Perspective on Forecasting and Integration of Renewables in Power Systems, Technical Report, Danish Energy Agency, 2020.

[7] G. Kariniotakis, S. Camal, Smart4RES team, Smart4RES: Improved weather modelling and forecasting dedicated to renewable energy applications, in: EGU General Assembly, online, 2021, p. 2. doi:10.5194/egusphere-egu21-16219.

[8] C. W. Hansen, W. F. Holmgren, A. Tuohy, J. Sharp, A. T. Lorenzo, L. J. Boeman, A. Golnas, The solar forecast arbiter: An open source evaluation framework for solar forecasting, in: 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC), IEEE, Chicago, IL, USA, 2019, pp. 2452–2457. doi:10.1109/PVSC40753.2019.8980713.

[9] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, H. Zareipour, Energy forecasting: A review and outlook, IEEE Open J. Power Energy (2020) 376–388.

[10] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. Martinez-de-Pison, F. Antonanzas-Torres, Review of photovoltaic power forecasting, Sol. Energy 136 (2016) 78–111.

[11] K. Doubleday, V. Van Scyoc Hernandez, B.-M. Hodge, Benchmark probabilistic solar forecasts: Characteristics and recommendations, Sol. Energy 206 (2020) 52–67.

[12] R. H. Inman, H. T. Pedro, C. F. Coimbra, Solar forecasting methods for renewable energy integration, Prog. Energy Combust. Sci. 39 (2013) 535–576.

[13] S. Sobri, S. Koohi-Kamali, N. A. Rahim, Solar photovoltaic generation forecasting methods: A review, Energy Convers. Manag. 156 (2018) 459–497.

[14] R. Ahmed, V. Sreeram, Y. Mishra, M. Arif, A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization, Renew. Sust. Energy Rev. 124 (2020) 109792.

[15] D. Yang, J. Kleissl, C. A. Gueymard, H. T. Pedro, C. F. Coimbra, History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining, Sol. Energy 168 (2018) 60–101.

[16] J. Jung, R. P. Broadwater, Current status and future advances for wind speed and power forecasting, Renew. Sust. Energy Rev. 31 (2014) 762–777.

[17] G. Giebel, G. Kariniotakis, Wind power forecasting—a review of the state of the art, in: Renewable Energy Forecasting, Elsevier, 2017, pp. 59–109. doi:10.1016/B978-0-08-100504-0.00003-2.

[18] A. M. Foley, P. G. Leahy, A. Marvuglia, E. J. McKeogh, Current methods and advances in forecasting of wind power generation, Renew. Energy 37 (2012) 1–8.

[19] A. Tascikaraoglu, M. Uzunoglu, A review of combined approaches for prediction of short-term wind speed and power, Renew. Sust. Energy Rev. 34 (2014) 243–254.

[20] M. Santhosh, C. Venkaiah, D. M. Vinod Kumar, Current advances and approaches in wind speed and wind power forecasting for improved renewable energy integration: A review, Eng. Rep. 2 (2020).

[21] C. Sweeney, R. J. Bessa, J. Browell, P. Pinson, The future of forecasting for renewable energy, WIREs Energy Environ. (2019).

[22] F. Barbieri, S. Rajakaruna, A. Ghosh, Very short term photovoltaic power forecasting with cloud modeling-a review, Renew. Sust. Energy Rev. 75 (2017).

[23] M. Guermoui, F. Melgani, K. Gairaa, M. L. Mekhalfi, A comprehensive review of hybrid models for solar radiation forecasting, J. Clean. Prod. 258 (2020) 120357.

[24] T. Jensen, T. Fowler, B. Brown, J. Lazo, S. E. Haupt, Metrics for Evaluation of Solar Energy Forecasts, Technical Report NCAR/TN-527+STR, NCAR, 2016.

[25] Y. Zhang, J. Wang, X. Wang, Review on probabilistic forecasting of wind power generation, Renew. Sust. Energy Rev. 32 (2014) 255–270.

[26] M. S. Santos, R. C. Pereira, A. F. Costa, J. P. Soares, J. Santos, P. H. Abreu, Generating synthetic missing data: A review by missing mechanism, IEEE Access (2019) 11651–11667.

[27] P. Lauret, M. David, P. Pinson, Verification of solar irradiance probabilistic forecasts, Sol. Energy 194 (2019) 254–271.

[28] U. Government, Solar Voltaics Deployment May 2020, Accessed 28/07/2020.
[29] U. Government, Renewable Energy Planning Database, Accessed 28/07/2020.
[30] I. Neher, T. Buchmann, S. Crewell, B. Pospichal, S. Meilinger, Impact of atmospheric aerosols on solar power, Meteorol. Z. 28 (2019) 305–321.
[31] B. Stridh, Evaluation of economical benefit of cleaning of soiling and snow in PV plants at three European locations, in: 2012 38th IEEE Photovoltaic Specialists Conference, IEEE, Austin, TX, USA, 2012, pp. 1448–1451. doi:10.1109/PVSC.2012.6317869.
[32] H. T. Pedro, R. H. Inman, C. F. Coimbra, Mathematical methods for optimized solar forecasting, in: Renewable Energy Forecasting, Elsevier, 2017, pp. 111–152. doi:10.1016/B978-0-08-100504-0.00004-4.
[33] H. Yang, B. Kurtz, D. Nguyen, B. Urquhart, C. W. Chow, M. Ghonima, J. Kleissl, Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego, Sol. Energy 103 (2014) 502–524.
[34] F. Wang, Z. Zhen, C. Liu, Z. Mi, B.-M. Hodge, M. Shafie-khah, J. P. Catalão, Image phase shift invariance based cloud motion displacement vector calculation method for ultra-short-term solar PV power forecasting, Energy Convers. Manag. 157 (2018) 123–135.
[35] T. Schmidt, J. Kalisch, E. Lorenz, D. Heinemann, Evaluating the spatio-temporal performance of sky-imager-based solar irradiance analysis and forecasts, Atmos. Chem. Phys. 16 (2016) 3399–3412.
[36] J. Lago, K. De Brabandere, F. De Ridder, B. De Schutter, Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data, Sol. Energy 173 (2018) 566–577.
[37] T. M. Harty, W. F. Holmgren, A. T. Lorenzo, M. Morzfeld, Intra-hour cloud index forecasting with data assimilation, Sol. Energy 185 (2019) 270–282.
[38] K. Bellinguer, R. Girard, G. Bontron, G. Kariniotakis, Short-term forecasting of photovoltaic generation based on conditioned learning of geopotential fields, in: 2020 55th International Universities Power Engineering Conference (UPEC), IEEE, Torino, Italy, 2020, pp. 1–6. doi:10.1109/UPEC49904.2020.9209858.
[39] T. Carriere, C. Vernay, S. Pitaval, G. Kariniotakis, A novel approach for seamless probabilistic photovoltaic power forecasting covering multiple time frames, IEEE Trans. Smart Grid 11 (2020) 2281–2292.
[40] M. Fliess, C. Join, C. Voyant, Prediction bands for solar energy: New short-term time series forecasting techniques, Sol. Energy 166 (2018) 519–528.
[41] H. Lee, N.-W. Kim, J.-G. Lee, B.-T. Lee, Uncertainty-aware forecast interval for hourly PV power output, IET Renew. Power Gener. 13 (2019) 2656–2664.
[42] F. Golestaneh, P. Pinson, H. B. Gooi, Very short-term nonparametric probabilistic forecasting of renewable energy generation— with application to solar energy, IEEE Trans. Power Syst. 31 (2016) 3850–3863.
[43] H. Sheng, J. Xiao, Y. Cheng, Q. Ni, S. Wang, Short-term solar power forecasting based on weighted gaussian process regression, IEEE Trans. Ind. Electron. 65 (2018) 300–308.
[44] I. Majumder, P. Dash, R. Bisoi, Variational mode decomposition based low rank robust kernel extreme learning machine for solar irradiation forecasting, Energy Convers. Manag. 171 (2018) 787–806.
[45] C. Huang, L. Wang, L. L. Lai, Data-driven short-term solar irradiance forecasting based on information of neighboring sites, IEEE Trans. Ind. Electron. 66 (2019) 9918–9927.
[46] M. Rana, I. Koprinska, V. G. Agelidis, Univariate and multivariate methods for very short-term solar photovoltaic power forecasting, Energy Convers. Manag. 121 (2016) 380–390.
[47] B. Sivaneasan, C. Yu, K. Goh, Solar forecasting using ANN with fuzzy logic pre-processing, Energy Procedia 143 (2017) 727–732.
[48] X. Luo, J. Sun, L. Wang, W. Wang, W. Zhao, J. Wu, J.-H. Wang, Z. Zhang, Short-term wind speed forecasting via stacked extreme learning machine with generalized correntropy, IEEE Trans. Ind. Inf. 14 (2018) 4963–4971.
[49] P. Tang, D. Chen, Y. Hou, Entropy method combined with extreme learning machine method for the short-term photovoltaic power generation forecasting, Chaos Soliton Fract. 89 (2016) 243–248.
[50] M. Abuella, B. Chowdhury, Random forest ensemble of support vector regression models for solar power forecasting, in: 2017 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), IEEE, North America, 2017, pp. 1–5.
[51] A. T. Eseye, J. Zhang, D. Zheng, Short-term photovoltaic solar power forecasting using a hybrid wavelet-PSO-SVM model based on SCADA and meteorological information, Renew. Energy 118 (2018) 357–367.
[52] C. Li, Z. Xiao, X. Xia, W. Zou, C. Zhang, A hybrid model based on synchronous optimisation for multi-step short-term wind speed forecasting, Appl. Energy 215 (2018) 131–144.
[53] X. G. Agoua, R. Girard, G. Kariniotakis, Short-term spatio-temporal forecasting of photovoltaic power production, IEEE Trans. Sustain. Energy 9 (2018) 538–546.
[54] S. E. Haupt, B. Kosović, T. Jensen, J. K. Lazo, J. A. Lee, P. A. Jiménez, J. Cowie, G. Wiener, T. C. McCandless, M. Rogers, S. Miller, M. Sengupta, Y. Xie, L. Hinkelman, P. Kalb, J. Heiser, Building the Sun4Cast system: Improvements in solar power forecasting, Bull. Amer. Meteor. Soc. 99 (2018) 121–136.
[55] J. A. Lee, S. E. Haupt, P. A. Jiménez, M. A. Rogers, S. D. Miller, T. C. McCandless, Solar irradiance nowcasting case studies near Sacramento, J. Appl. Meteorol. Climatol. 56 (2017) 85–108.
[56] P. Pinson, Wind energy: Forecasting challenges for its operational management, Stat. Sci. 28 (2013) 564–585.
[57] M. Méchali, R. Barthelmie, S. Frandsen, L. Jensen, P.-E. Réthoré, Wake effects at Horns Rev and their influence on energy production, in: Proc. European Wind Energy Conference, 2006, p. 10.
[58] F. Lamraoui, G. Fortin, R. Benoit, J. Perron, C. Masson, Atmospheric icing impact on wind turbine production, Cold Reg. Sci. Technol. 100 (2014) 36–49.
[59] W. Michael, Long term performance of wind farms, http://www.ewea.org/events/workshops/wp-content/uploads/2014/12/Tech14a2-1-Wilkinson.pdf, Accessed 09/11/2020.
[60] R. Tawn, J. Browell, I. Dinwoodie, Missing data in wind farm time series: Properties and effect on forecasts, Electr. Power Syst. Res. 189

(2020) 106640.
[61] H. Zhou, Z. Wang, A multiple-model based adaptive control algorithm for very-short term wind power forecasting, in: 2016 IEEE International Conference on Power System Technology (POWERCON), IEEE, Wollongong, Australia, 2016, pp. 1–6. doi:10.1109/POWERCON.2016.7753882.
[62] L. Cavalcante, R. J. Bessa, M. Reis, J. Browell, LASSO vector autoregression structures for very short-term wind power forecasting, Wind Energy 20 (2017) 657–675.
[63] J. W. Messner, P. Pinson, Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting, Int. J. Forecast. 35 (2019) 1485–1498.
[64] J. Dowell, P. Pinson, Very-short-term probabilistic wind power forecasts by sparse vector autoregression, IEEE Trans. Smart Grid (2015) 763–770.
[65] Y. Zhao, L. Ye, P. Pinson, Y. Tang, P. Lu, Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting, IEEE Trans. Power Syst. 33 (2018) 5029–5040.
[66] J. Browell, D. R. Drew, K. Philippopoulos, Improved very short-term spatio-temporal wind forecasting using atmospheric regimes, Wind Energy (2018) 968–979.
[67] C. Gonçalves, R. J. Bessa, P. Pinson, A critical overview of privacy-preserving approaches for collaborative forecasting, Int. J. Forecast. 37 (2021) 322–342.
[68] Y. Zhang, J. Wang, A distributed approach for wind power probabilistic forecasting considering spatio-temporal correlation without direct access to off-site information, IEEE Trans. Power Syst. 33 (2018) 5714–5726.
[69] A. Chaudhary, A. Sharma, A. Kumar, K. Dikshit, N. Kumar, Short term wind power forecasting using machine learning techniques, J Stat. Manag. Syst. 23 (2020) 145–156.
[70] M. Landry, Probabilistic gradient boosting machines for GEFCom2014 wind forecasting, Int. J. Forecast. (2016) 1061–1066.
[71] S. Jiang, R. Fang, L. Wang, C. Peng, Very short-term wind power forecasting based on SVM-Markov, in: Proc. Int. Conference Advances Energy Environ. Chem. Eng., Atlantis Press, Changska, China, 2015, pp. 130–134. doi:10.2991/aeece-15.2015.27.
[72] J. He, J. Xu, Ultra-short-term wind speed forecasting based on support vector machine with combined kernel function and similar data, J Wireless Com. Network 2019 (2019).
[73] J. Wang, J. Heng, L. Xiao, C. Wang, Research and application of a combined model based on multi-objective optimization for multi-step ahead wind speed forecasting, Energy 125 (2017) 591–613.
[74] P. Du, J. Wang, Z. Guo, W. Yang, Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting, Energy Convers. Manag. 150 (2017) 90–107.
[75] P. Du, J. Wang, W. Yang, T. Niu, A novel hybrid model for short-term wind power forecasting, Appl. Soft Comput. 80 (2019) 93–106.
[76] M. Khodayar, O. Kaynak, M. E. Khodayar, Rough deep neural architecture for short-term wind speed forecasting, IEEE Trans. Ind. Inf. 13 (2017) 2770–2779.
[77] NREL, Western Wind data set, https://www.nrel.gov/grid/western-wind-data.html, Accessed 2021-02-04.
[78] F. Rodríguez, A. M. Florez-Tapia, L. Fontán, A. Galarza, Very short-term wind power density forecasting through artificial neural networks for microgrid control, Renew. Energy 145 (2020) 1517–1527.
[79] M. Khodayar, J. Wang, Spatio-temporal graph deep neural network for short-term wind speed forecasting, IEEE Trans. Sustain. Energy 10 (2019) 670–681.
[80] M. A. Hossain, R. K. Chakrabortty, S. Elsawah, M. J. Ryan, Very short-term forecasting of wind power generation using hybrid deep learning model, Journal of Cleaner Production 296 (2021) 126564.
[81] X. Ma, Y. Jin, Q. Dong, A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting, Appl. Soft Comput. 54 (2017) 296–312.
[82] Z. Zhang, H. Qin, Y. Liu, Y. Wang, L. Yao, Q. Li, J. Li, S. Pei, Long short-term memory network based on neighborhood gates for processing complex causality in wind speed prediction, Energy Convers. Manag. 192 (2019) 37–51.
[83] J. Chen, G.-Q. Zeng, W. Zhou, W. Du, K.-D. Lu, Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization, Energy Convers. Manag. 165 (2018) 681–695.
[84] H. Liu, X.-W. Mi, Y.-F. Li, Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network, Energy Convers. Manag. 156 (2018) 498–514.
[85] P. Lu, L. Ye, B. Sun, C. Zhang, Y. Zhao, J. Teng, A new hybrid prediction method of ultra-short-term wind power forecasting based on EEMD-PE and LSSVM optimized by the GSA, Energies 11 (2018) 697.
[86] J. Zhang, Y. Wei, Z.-F. Tan, W. Ke, W. Tian, A hybrid method for short-term wind speed forecasting, Sustainability 9 (2017) 596.
[87] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, H. H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. A 454 (1996) 903–995.
[88] Z. Liu, P. Jiang, L. Zhang, X. Niu, A combined forecasting model for time series: Application to short-term wind speed forecasting, Appl. Energy 259 (2020) 114137.
[89] W. Zhang, Z. Qu, K. Zhang, W. Mao, Y. Ma, X. Fan, A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting, Energy Convers. Manag. 136 (2017) 439–451.
[90] Aasim, S. Singh, A. Mohapatra, Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting, Renew. Energy 136 (2019) 758–768.
[91] C. Zhang, J. Zhou, C. Li, W. Fu, T. Peng, A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting, Energy Convers. Manag. 143 (2017) 360–376.
[92] S. G. Hall, J. Mitchell, Combining density forecasts, Int. J. Forecast. 23 (2007) 1–13.
[93] F. X. Diebold, Forecast combination and encompassing: Reconciling two divergent literatures, Int. J. Forecast. 5 (1989) 589–592.
[94] J. Shi, Z. Ding, W.-J. Lee, Y. Yang, Y. Liu, M. Zhang, 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 (2014) 521–526.

[95] L. Xiao, J. Wang, Y. Dong, J. Wu, Combined forecasting models for wind energy forecasting: A case study in China, Renew. Sust. Energy Rev. 44 (2015) 271–288.

[96] A. Zameer, J. Arshad, A. Khan, M. A. Z. Raja, Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks, Energy Convers. Manag. 134 (2017) 361–372.

[97] T. Ouyang, X. Zha, L. Qin, A combined multivariate model for wind power prediction, Energy Convers. Manag. 144 (2017) 361–373.

[98] Y. Lin, M. Yang, C. Wan, J. Wang, Y. Song, A multi-model combination approach for probabilistic wind power forecasting, IEEE Trans. Sustain. Energy 10 (2019) 226–237.

[99] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, R. J. Hyndman, Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond, Int. J. Forecast. 32 (2016) 896–913.

[100] C. Feng, M. Cui, B.-M. Hodge, J. Zhang, A data-driven multi-model methodology with deep feature selection for short-term wind forecasting, Appl. Energy 190 (2017) 1245–1257.

[101] P. Li, X. Wang, J. Yang, Short-term wind power forecasting based on two-stage attention mechanism, IET Renew. Power Gener. 14 (2020) 297–304.

[102] P. Jiang, R. Li, H. Li, Multi-objective algorithm for the design of prediction intervals for wind power forecasting model, Appl. Math. Model. 67 (2019) 101–122.

[103] H.-Z. Wang, G.-Q. Li, G.-B. Wang, J.-C. Peng, H. Jiang, Y.-T. Liu, Deep learning based ensemble approach for probabilistic wind power forecasting, Appl. Energy 188 (2017) 56–70.

[104] A. Carpinone, M. Giorgio, R. Langella, A. Testa, Markov chain modeling for very-short-term wind power forecasting, Electr. Power Syst. Res. 122 (2015) 152–158.

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

[106] Z. Zhang, L. Ye, H. Qin, Y. Liu, C. Wang, X. Yu, X. Yin, J. Li, Wind speed prediction method using shared weight long short-term memory network and Gaussian process regression, Appl. Energy 247 (2019) 270–284.

[107] C. Gilbert, J. Browell, D. McMillan, Leveraging turbine-level data for improved probabilistic wind power forecasting, IEEE Trans. Sustain. Energy 11 (2020) 1152–1160.

[108] P. Jiang, Y. Wang, J. Wang, Short-term wind speed forecasting using a hybrid model, Energy 119 (2017) 561–577.

[109] A. A. Ezzat, Turbine-specific short-term wind speed forecasting considering within-farm wind field dependencies and fluctuations, Appl. Energy 269 (2020) 115034.

[110] Y. Ding, Data Science for Wind Energy, first ed., Chapman and Hall/CRC, 2019.

[111] R. Yu, J. Gao, M. Yu, W. Lu, T. Xu, M. Zhao, J. Zhang, R. Zhang, Z. Zhang, LSTM-EFG for wind power forecasting based on sequential correlation features, Future Gener. Comput. Syst. 93 (2019) 33–42.

[112] I. Würth, L. Valldecabres, E. Simon, C. Möhrlen, B. Uzunoğlu, C. Gilbert, G. Giebel, D. Schlipf, A. Kaifel, Minute-Scale Forecasting of Wind Power—Results from the Collaborative Workshop of IEA Wind Task 32 and 36, Energies 12 (2019) 712.

[113] L. Valldecabres, N. Nygaard, L. Vera-Tudela, L. von Bremen, M. Kühn, On the use of dual-doppler radar measurements for very short-term wind power forecasts, Remote Sens. 10 (2018) 1701.

[114] L. Valldecabres, A. Peña, M. Courtney, L. von Bremen, M. Kühn, Very short-term forecast of near-coastal flow using scanning lidars, Wind Energy Sci. 3 (2018) 313–327.

[115] S. G. Benjamin, S. S. Weygandt, J. M. Brown, M. Hu, C. R. Alexander, T. G. Smirnova, J. B. Olson, E. P. James, D. C. Dowell, G. A. Grell, H. Lin, S. E. Peckham, T. L. Smith, W. R. Moninger, J. S. Kenyon, G. S. Manikin, A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh, Mon. Weather Rev. 144 (2016) 1669–1694.

[116] M. Milan, B. Macpherson, R. Tubbs, G. Dow, G. Inverarity, M. Mittermaier, G. Halloran, G. Kelly, D. Li, A. Maycock, T. Payne, C. Piccolo, L. Stewart, M. Wlasak, Hourly 4D-Var in the Met Office UKV operational forecast model, Q. J. R. Meteorol. Soc. 146 (2020) 1281–1301.

[117] S. Hagelin, J. Son, R. Swinbank, A. McCabe, N. Roberts, W. Tennant, The Met Office convective-scale ensemble, MOGREPS-UK, Q. J. R. Meteorol. Soc. 143 (2017) 2846–2861.

[118] J. R. Lawson, J. S. Kain, N. Yussouf, D. C. Dowell, D. M. Wheatley, K. H. Knopfmeier, T. A. Jones, Advancing from Convection-Allowing NWP to Warn-on-Forecast: Evidence of Progress, Weather Forecast. 33 (2018) 599–607.

[119] J. W. Messner, P. Pinson, J. Browell, M. B. Bjerregård, I. Schicker, Evaluation of wind power forecasts—An up-to-date view, Wind Energy 23 (2020) 1461–1481.

[120] M. E. Torres, M. A. Colominas, G. Schlotthauer, P. Flandrin, A complete ensemble empirical mode decomposition with adaptive noise, in: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Prague, Czech Republic, 2011, pp. 4144–4147. doi:10.1109/ICASSP.2011.5947265.

[121] P. Chen, K. K. Berthelsen, B. Bak-Jensen, Z. Chen, Markov model of wind power time series using Bayesian inference of transition matrix, in: 2009 35th Annual Conference of IEEE Industrial Electronics, IEEE, Porto, Portugal, 2009, pp. 627–632. doi:10.1109/IECON.2009.5414993.

[122] C. Möhrlen, J. Zack, Recommended Practices for Selecting Renewable Power Forecasting Solutions, Technical Report, IEA Wind Task 36, 2019.

[123] B. Efron, Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods, Biometrika 68 (1981) 589–599.

[124] F. X. Diebold, R. S. Mariano, Comparing predictive accuracy, J Bus. Econ. Stat. 13 (1995) 253–263.

[125] P. Pinson, Introducing distributed learning approaches in wind power forecasting, in: 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), IEEE, Beijing, 2016, pp. 1–6. doi:10.1109/PMAPS.2016.7764224.

[126] C. Goncalves, P. Pinson, R. J. Bessa, Towards data markets in renewable energy forecasting, IEEE Trans. Sustain. Energy 12 (2021) 533–542.

[127] G. Chang, H. Lu, Y. Chang, Y. Lee, An improved neural network-based approach for short-term wind speed and power forecast, Renew.

Energy 105 (2017) 301–311.
[128] A. Dupré, P. Drobinski, B. Alonzo, J. Badosa, C. Briard, R. Plougonven, Sub-hourly forecasting of wind speed and wind energy, Renew. Energy 145 (2020) 2373–2379.
[129] W. Fu, K. Wang, J. Tan, K. Zhang, A composite framework coupling multiple feature selection, compound prediction models and novel hybrid swarm optimizer-based synchronization optimization strategy for multi-step ahead short-term wind speed forecasting, Energy Convers. Manag. 205 (2020) 112461.
[130] Y. Noorollahi, M. A. Jokar, A. Kalhor, Using artificial neural networks for temporal and spatial wind speed forecasting in Iran, Energy Convers. Manag. 115 (2016) 17–25.
[131] L. Ye, Y. Zhao, C. Zeng, C. Zhang, Short-term wind power prediction based on spatial model, Renew. Energy 101 (2017) 1067–1074.

Appendix

Table A1: Table of papers reviewed and broad categorisation of types of methods used.

Paper	Horizon	Solar	Wind	Probabilistic	ML	Statistical	Decomposition	hybrid/combination	turbine-level	image based
Abuella (2017)[50]	day ahead	✓			✓			✓		
Agoua (2018) [53]	0-6hrs	✓				✓				
Bellinguer (2020) [38]	0-6hrs	✓				✓				✓
Carriere (2020) [39]		✓								✓
Eseye (2018) [51]	3-24 hrs	✓			✓		✓	✓		
Fliess (2019) [40]	1,15,60 min	✓		✓						
Golestaneh (2016) [42]		✓		✓	✓					
Harty (2019) [37]	15-60min	✓						✓		✓
Huang (2019) [45]	30-120min	✓			✓					
Lago (2018) [36]		✓			✓					✓
Lee (2016) [55]	15min-6hrs	✓		✓	✓			✓		✓
Lee (2019) [41]		✓		✓	✓					
Li (2018) [52]		✓			✓		✓			
Luo (2018) [48]	5secs & 10-60min ahead	✓			✓					
Majumder (2018) [44]	15min - 1 day	✓			✓		✓			
Rana (2016) [46]	5-60min	✓			✓					
Schmidt (2016) [35]	up to 25min	✓			✓					✓
Sheng (2018) [43]	5min	✓			✓					
Sivaneasan (2017) [47]	minutes	✓			✓					
Tang (2016) [49]		✓			✓					
Wang (2018) [34]		✓								✓
Aasim (2019) [90]	10 minutes		✓			✓	✓			
Bracale (2015) [105]	up to 24hrs		✓	✓		✓				
Browell (2018) [66]	1-6hrs		✓			✓				
Carpinone (2015) [104]	10min		✓	✓						
Cavalcante (2017) [62]			✓			✓				
Chang (2017) [127]			✓		✓					
Chaudhary (2020) [69]			✓		✓					
Chen (2018) [83]	10min, 1hr		✓		✓	✓				
Dowell (2016) [64]	5 min		✓			✓				
Du (2017) [74]	10,30,60min		✓		✓					
Du (2019) [75]			✓		✓		✓			

Paper	Horizon	Solar	Wind	Probabilistic	ML	Statistical	Decomposition	hybrid/combination	turbine-level	image based
Dupre (2020) [128]	10-170mins		✓			✓				
Ezzat (2020) [109]	1-12hrs		✓	✓		✓			✓	
Feng (2017) [100]	1hr		✓	✓	✓			✓		
Fu (2020) [129]	10min - 1hr		✓		✓		✓	✓		
Gilbert (2020) [107]	0-48hrs		✓	✓	✓				✓	
He (2019) [72]			✓		✓					
Hong (2016) [99]			✓	✓						
Hossain (2021) [80]	10-45min		✓		✓					
Jiang (2015) [71]	10min		✓		✓					
Jiang (2017) [108]			✓		✓					
Jiang (2019) [102]	10,30,60mins		✓	✓	✓			✓		
Khodayar (2017) [76]	10min-3hrs		✓		✓					
Khodayar (2019) [79]	1-24hrs		✓		✓					
Lin (2019) [98]	up to 24hrs		✓	✓		✓		✓		
Li (2020) [101]			✓	✓	✓					
Liu (2020) [88]	10-30mins		✓	✓	✓	✓	✓	✓		
Liu (2018) [84]			✓		✓		✓	✓		
Lu (2018) [85]	15min-1hr		✓		✓		✓			
Ma (2017) [81]	10,30,60mins		✓		✓					
Messner (2019) [63]	up to 1hr		✓			✓				
Noorollahi (2016) [130]			✓							
Ouyang (2017) [97]	1-10hr		✓					✓		
Rodriguez (2020) [78]	10min		✓	✓	✓					
Shi (2014) [94]	15min		✓		✓			✓		
Valledecabres (2018) [113]	5min		✓	✓					✓	✓
Wang (2017) [103]	1 hr		✓	✓	✓		✓			
Wang (2017) [73]	10-30mins		✓		✓			✓		
Wurth (2019) [112]	1hr		✓						✓	
Xiao (2015) [95]			✓					✓		
Ye (2017) [131]			✓							
Yu (2019) [111]	90mins		✓		✓				✓	
Zameer (2017) [96]	1hr		✓		✓			✓		
Zhang (2017) [89]			✓		✓		✓	✓		
Zhang (2017) [91]			✓		✓		✓			
Zhang (2017) [86]	3-24hrs		✓		✓	✓	✓	✓		
Zhang (2018) [68]	up to 24hrs		✓	✓		✓				
Zhang (2019) [82]			✓		✓					
Zhang (2019) [106]			✓	✓	✓					
Zhao (2018) [65]	up to 1.5hrs		✓			✓				
Zhou (2016) [61]	15min - 4 hrs		✓			✓		✓		
totals (% of studies)		29%	71%	26%	64%	24%	18%	25%	7%	11%