A transfer strategy for power output estimation of wind farm at planning stage based on SVR model

A transfer strategy for power output estimation of wind farm at planning stage based on SVR model

Zihao Li, Wei Sun, *Member, IEEE, Yue Xiang, Senior Member, IEEE,* Gareth P. Harrison, *Senior Member, IEEE*

***Abstract*— The conventional wind farm (WF) power generation modelling method highly relies on wind hindcast produced by record time-series data or numerical weather modelling. However, estimating production at future sites is challenging in the absence of local wind monitoring. To address this, a data-driven WF modelling and model transfer strategy is proposed in this work. It considers the challenge of how to transpose metered data from existing operational WFs to sites that might feature as a prospective site for a new WF. By modelling 14 WFs distributed across Scotland using a machine learning (ML) approach, this study proved that it was possible to effectively model metered production at a site using modelled wind speed and direction. In addition, this study also found that when the latitude difference between two WFs is less than 0.2 degrees and the distance is less than 50km, two WFs in non-mountainous areas can share an ML model. The results of the shared ML model remain superior to the results of the given power curve from manufacturers, after adjusting the results by the ratio of the power curve in these two WFs. The WF model transfer strategy investigated in this work offered a novel approach to transposing WF production estimates to new sites and appeared to offer better value than simple power curves, which is of importance at the early planning stage for sites selection, although it would likely not fully replace detailed micro-siting modelling as are well established in the industry.**

***Index Terms*— wind farm, power output estimation, machine learning, model transfer strategy, power curve.**

I. INTRODUCTION

Facing the challenge of climate change, the UK government has committed to reduce the carbon emissions to net zero by 2050 [1]. This is an ambitious target for this nation, which for decades has relied on carbon-based sources for most of its electricity supply, such as coal, oil and gas [2]. Although this country has realized this problem and has shown significant progress in deploying renewable energy to transit away from fossil fuels, it still requires up to another 160 GW of wind energy to meet the 2050 deadline [3,4]. In order to facilitate the development of wind energy, i.e., determine wind farm (WF)'s capacity and its location properly, it is important to improve the accuracy of the planning stage WF power output estimates.

Present studies have designed power generation models using open-access data from WFs [5,6]. The basic and the earliest WF power output model is achieved by 'wind turbine (WT) power curve scale up'. According to the power curves of WTs specified by the manufacturers, the whole WF is assumed to be a single WT and its generation equals a single WT power output multiplied by the quantity of WT [7,8]. Such a method is used to estimate the production from a particular wind site during the early planning phase and can also be used when there are no other available WF output data.

However, as the capacity of WF grows, this basic model fails to capture the effect of site-specific factors on the power generation, such as the effects of wear and tear, ageing, terrain induced wind speed/direction variations and the presence of turbulence [9]. Estimates of wind power production involve uncertainties because of the stochastic nature of the wind and variation of the power curve [10]. Not considering these effects typically results in overestimating energy outputs. Nevertheless, it is a considerable challenge to describe these factors precisely in mathematical functions and investigate useful models. As a result, scientists start seeking help from the data-driven approach, which regards the WF as a black box and analyses the operational data in specific WF to study the statistical relationship among wind speed, wind direction and the power generation. During the last decade, machine learning (ML) is the most common technique to establish such a model [11,12], that can successfully avoid a complicated single WT modelling process. For example, paper [13] proposed three different machine learning models to build an equivalent steady state model of a WF. Reference [14] addressed an improved neural network model, which was suitable for ultra-short-term (10 min) and short-term wind speed forecasting. Yin and Zhao [15] explored the big data-driven multi-objective predictions for offshore WF based on machine learning.

Nevertheless, all these models share a common

This work was funded in part by EPSRC through the National Centre for Energy Systems Integration (grant number EP/P001173/1). (Corresponding author: Gareth Harrison.)

Z. Li, W. Sun and G. P. Harrison are with the School of Engineering, University of Edinburgh, Edinburgh, EH9 3DW, UK (e-mail: zihao.li@ed.ac.uk, w.sun@ed.ac.uk, gareth.harrison@ed.ac.uk).

Yue Xiang is with the College of Electrical Engineering, Sichuan University, Chengdu 610065, China (e-mail: xiang@scu.edu.cn).

Digital Object Identifier

disadvantage: they need to take the actual power output of the WF as part of the model input. Such a model cannot work for WFs without recorded power generation, for instance, at the planning stage. Planning stage WF power output estimation relies on a mixture of short-term onsite measurement, long time series of simulated wind speeds from mesoscale models and farm level modelling [16]. For many academic and industrial studies looking at, for example, system impacts, this is further reduced to a combination of mesoscale or reanalysis wind speed data and a power curve from a typical turbine or some form of generic aggregate power curve that intended to represent some of the farm level effects [17], which often introduces additional errors into the analysis. How to efficiently and accurately reflect farm level effects into a model without complex mathematical formulations is a huge challenge. In this regard, some existing studies provide valuable inspiration.

It is known that wind speed between geographically close WFs is not statistically independent but reflects spatial correlations, which refers to a relationship between the wind speed in one location and the wind speed in other locations around the geographic space. Wind speed correlation includes both temporal correlation within the same spatial position and spatial correlation between different spatial locations [18]. The wind speed value and power at a specific location are not only auto-correlated in terms of time, but also influenced by other factors related to spatial locations [19]. These factors include wake effect, topography, and roughness of the WF, pressure and temperature, etc. [20-22]. Therefore, depending on these characteristics of wind speed, appropriate spatial correlation models can be established between WFs at known positions. The model can infer the unknown wind speed and power value of adjacent locations from the wind speed and power value of several known WFs. For instance, publications [23-25] used neural network to demonstrate a spatial correlation wind power prediction model for two sites from 800m to 40km away. Using data in [26,27] analysed the high-frequency wind power output data in two WFs in North America, while paper [28] presented the spatio-temporal structure of wind power prediction errors in Denmark. Both works show a high nonlinear correlation of power outputs between WFs. Except for spatial correlation research among WFs, some other works try to use one or several WFs outputs to predict the whole area WF generations. Studies [29,30] found the wind power forecasting error for an area could be less than that for a single WF due to the spatial smoothing effect. The extent of error reduction depends on the size of the selected region and the number of WFs in the region.

Conclusions from these papers are sufficient to prove that it is evident that spatial correlation of WFs outputs exists under certain conditions. However, regardless of the methodology, it is worth noting that these papers only considered WFs that are already built and under operation, with the actual power outputs available in both WFs as inputs. Nevertheless, these prior works have not explicitly applied the method of spatial correlation to the study of WF *planning*. To fill the gap, in this work we first investigate a data-driven model of an existing WF, then use it to estimate the power curve in a planned WF that spatial correlates with the existing operating one. Such a novel attempt and the findings it brings about, are the main contribution of this paper.

In order to more accurately estimate the output power of the planning stage WF, this work presents a novel model transfer strategy to better estimate the relationship between WF level production and wind speed. To the author's best knowledge, this is the first work that proposed such a framework, which explicitly advises the planning stage WF to estimate its power output by establishing an ML model of an existing WF, rather than solely relying on the manufacturer's power curves. Specifically, it is found that for two WFs that are not in a mountainous area, they can share one ML model on the condition that the latitude difference between them is less than 0.2 degrees and the distance between them is less than 50km. The power estimation results from this strategy remain superior to the results of the given power curve. This suggests that if a WF is built near an existing WF, the power output of the target WF can be estimated using a specific model of the existing WF.

The rest of the paper is organized as follows. Section II describes the methodology applied in the case study, including equivalent power curve modelling, support vector regression (SVR) algorithms and transferring strategy. Section III introduces the case study and presents the results; section IV discusses these and draws some conclusions.

II. METHODOLOGY

The model transfer strategy addressed in the section contains three steps as illustrated in Figure 1. Since the installed capacity is different between WFs, under the conditions of the same wind speed and direction, the corresponding power generation is not directly comparable. It is necessary to normalize all WF generations to the same level, e.g. [0,1], before investigating the transfer strategy. An SVR power generation model for each WF is then developed with two inputs, wind speed and wind direction. Finally, cross-testing every single model on every WF and adjusting its output by the ratio of power curves in two WFs.

A. Using capacity factor power curve to model WF output

The overall production of a WF is equivalent to the sum of every single WT output, and its performance is generally described by the manufacturer's power curve. The simplest method for WF modelling is to represent the entire WF as a single equivalent WT with the corresponding equivalent power curve. This method is called "simple aggregated power curve" (SAPC) method in this work. In the modelling process, because the WF installed capacities vary from each other, the ML model designed for one WF may not be able to transfer directly to other WFs. To solve the problem, this work sets the rated power as one in every WF, and scales down the single WT power curve to obtain the equivalent WF power curve. In other words, instead of modelling the real power output in each WF, this work aims to model the WF power generation capacity factor, CF_{WF}, which is defined as Eq.1:

$$CF_{\mathrm{WF}} = {P_{\mathrm{r}}}^{t} \Big/ {P_{\mathrm{pc}}}^{t} \tag{1}$$

The P_r^t is the real power generation in WF at time t; while the P_{pc}^t is the corresponding power curve generation at time t.

B. Machine learning to model WF output

Support Vector Regression (SVR): SVR is an extension of the Support Vector Machine [31]. The principle of SVR is to cluster data into different groups, making sure that the nearest points in one group have maximum distance with the classify boundary (hyperplane). In theory, any linear function $F(X)$ exists to formulate the non-linear relationship between the input and output data, such a function, namely the SVR function. The general function of SVR estimation:

$$f(\mathrm{x},\mathrm{w}) = \sum_{j=1}^{m} w_i g_i(x) + b \tag{2}$$

where x is the input variable, i.e., wind speed and wind direction in this paper, $\{g_i(x)\}_{j=1}^{m}$ are features, b and w_i are coefficients, (bias and weights of input variable x) that have to be estimated from the data. Secondly, we need to find a pair of minimal norm values $({w_i}'w_i)$. According to [32], this is formulated as a convex quadratic programming problem:

$$\min \ \frac{1}{2}||w_i||^2 \tag{3}$$

subject to:

$$\left|y_i - ({w_i}^{\mathrm{T}} x_i + b_i)\right| \le \epsilon \tag{4}$$

where y_i is the training sample target, i.e., in this paper, the WF power output, and ϵ denotes the desired error range for all points. However, it is sometimes impossible to find a function that satisfies these constraints for all points, therefore, the slack variables ξ_i and ξ_i^* are used to guarantee that a solution exists for all points. The objective function becomes:

$$\min \ \frac{1}{2}||w_i||^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^*) \tag{5}$$

subject to:

$$y_i - w^{\mathrm{T}}\phi(x_i) - b_i \le \epsilon + \xi_i \tag{6}$$

$$w^{\mathrm{T}}\phi(x_i) + b_i - y_i \le \epsilon + \xi_i^* \tag{7}$$

$$\xi_i, \xi_i^* \ \ge 0 \tag{8}$$

where C is the box constraint, a positive value that controls the penalty imposed on observations that lie outside the error margin (ϵ) [32]; and ϕ represents a kernel function for mapping the input space to a higher dimensional feature space. Reference [33] provides a detailed overview.

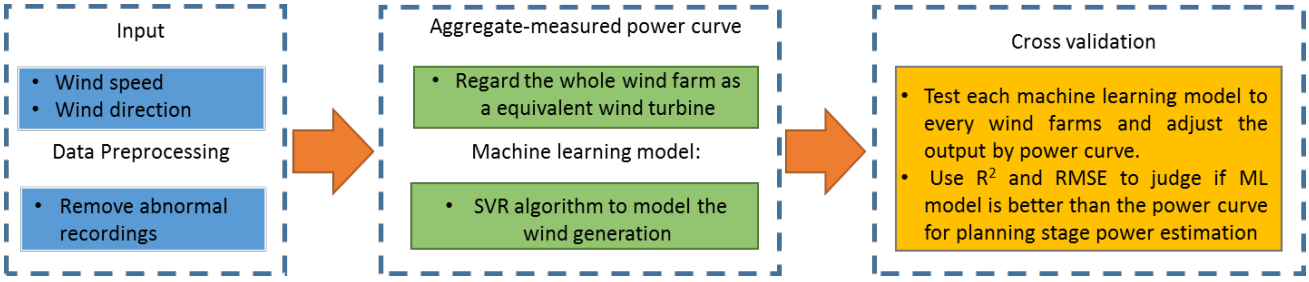

Figure 1. Transfer strategy working flow

C. Model cross validation

For WF that provides data for the model training process, the SVR model should be able to capture its power output precisely. However, the extent to which the model derived from a WF can estimate the power output of another WF that was not included during the model training process, is an interesting question.

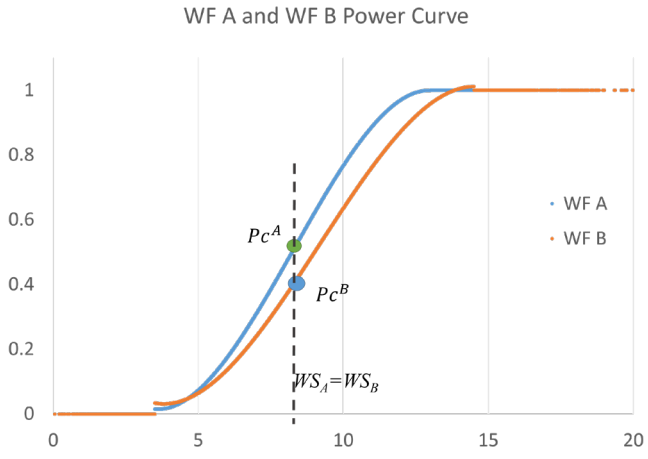

Figure 2. Graphical interpretation for Eq.9 and Eq .10

The power curves of all WTs are obtained in a standardized test environment, e.g., IEC 61400-12-1:2017 [34]. The output of the whole WF can be approximated by adding up the output of each WT using the power curve and wind speed. Therefore, for any two WFs A and B (for simplicity, assuming that a WF consists of identical WTs), the ratio of WF output A, P_A, and output B, P_B, equals the ratio of corresponding WT power output calculated from their power curves, Pc^A and Pc^B, if the wind speed at A and B, i.e., ws_A and ws_B, are the same. That is (Graphical interpretation in Figure 2.):

$$P_A/P_B = Pc^A/Pc^B \tag{9}$$

subject to:

$$ws_A = ws_B \tag{10}$$

In practice, the above formula may oversimplify the relationship between power generation and wind speed, since the other neglected factors, such as turbulence, terrain, and wind direction, etc., could also affect the output of the WF. Although it is evident that WF power output would not precisely follow the WT power curve in the real world, when it comes to estimating the output of WF that is still at the planning

stage, power curve of WT provided by manufacturers is often the only data available. Therefore, how to better use WT power curve to improve the estimation accuracy of planned WF outputs is a worthy topic of study. Considering that the WFs have spatial correlations under certain conditions, if we construct an ML model that simulate the power output of a WF successfully, is it still feasible to transfer the output model between different WFs based on the power curve?

In this work, taking the wind speed and wind direction of each WF as input, a series of ML models are designed for each WF. Then, these ML models are applied to all WFs in turn to obtain a set of 'dislocated results'; for example, applying the SVR model trained on WF A (WF B) to WF B (WF A) to obtain a result, *Rab (Rba)*. Next, this output will be adjusted by the power estimation results from the power curve, i.e., Pc^A/Pc^B, to scale up or down the *Rab (Rba)* to get the adjusted output, *'A_TO_B'('B_TO_A')*. Consequently, *'A_TO_B' ('B_TO_A')* is able to simulate wind power output in WF B (WF A) better than *Rab (Rba)*. Figure 3 illustrates the general steps.

It should be pointed out that the process of obtaining *'A_TO_B'* does not require the actual output of WF B. It only needs the WT information in WF A and B, wind speed and wind direction of WF A and B, and the actual power output of WF A. Normally, these pieces of information are already available when WF B is in the planning stage. Therefore, this strategy can be implemented to WF output simulation without additional requirements. Of course, as Figure 3 demonstrates, whether the performance obtained by this method is more accurate than SAPC is still to be analysed.

D. Performance and analysis.

The modelling results are evaluated with two metrics to assess the performance of different methods, root mean square error (RMSE) and coefficient of determination (R^2). RMSE is used to measure how concentrated the data is around the line of best fit, while the R^2 provides a measure of how well observed outcomes are replicated by the model. Lower RMSE and a higher R^2 indicate better model results. Eq.11 and Eq.12 are formulas for RMSE and R^2, where y_i^{Wind} is the real WF generation value while the $y_i^{'}$ is the simulated generation and n is the number of samples.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{\text{Wind}} - y_i^{'})^2} \tag{11}$$

$$\text{R}^2 = \left(\frac{E(yy_i^{\text{Wind}}) - E(y)E(y_i^{\text{Wind}})}{\sqrt{E(y_i^{'2}) - (E(y_i^{'}))^2}\sqrt{E(y_i^{\text{Wind}2}) - (E(y_i^{\text{Wind}}))^2}}\right)^2 \tag{12}$$

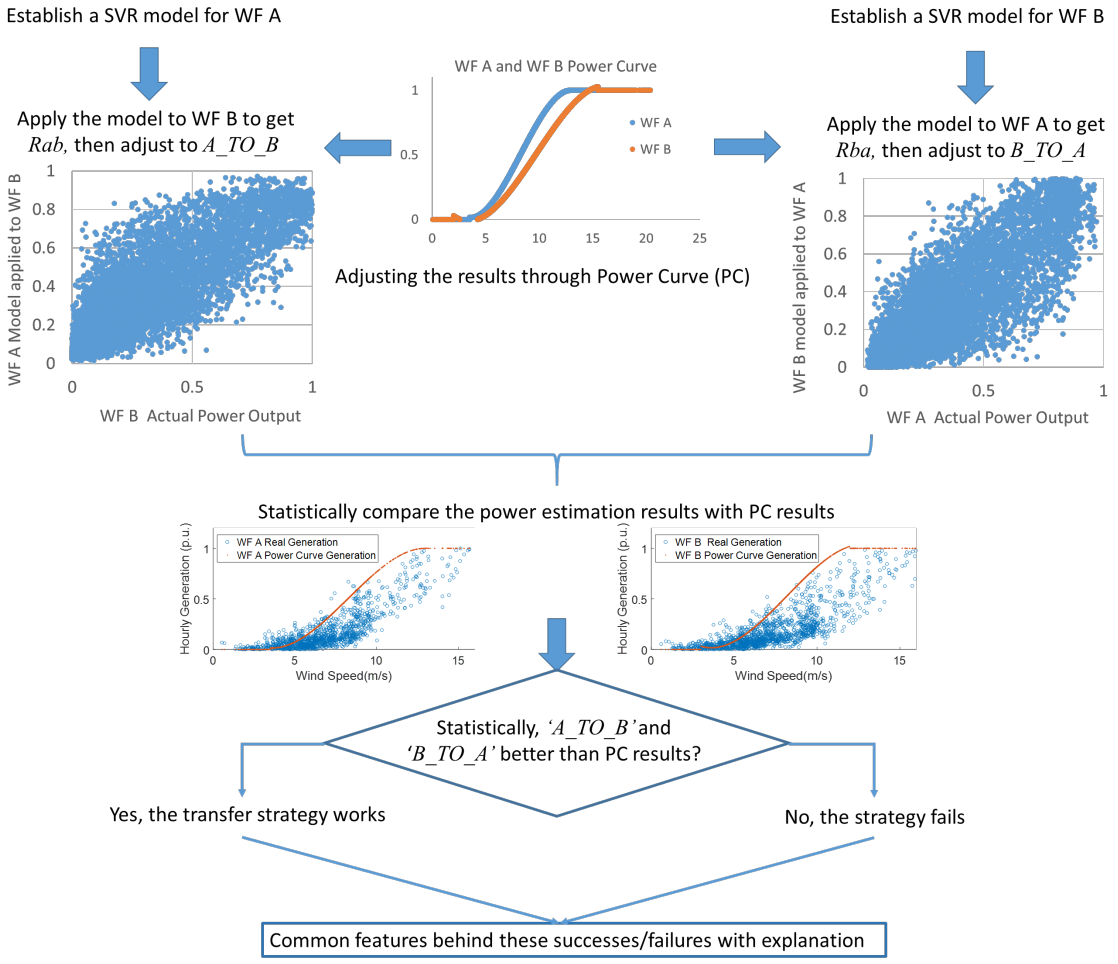

Figure3. Model cross testing process

III. CASE STUDY AND RESULTS

In this study, 14 WFs located in Scotland were selected to validate the proposed model transfer strategy. These WFs are spread across Scotland and use many different types of WTs. Simulated hindcast wind speed and wind direction data were collected as well as the actual hourly generation recordings of all the WFs from 2009 to 2010. Due to inconsistency in the timing of grid connection of WFs and other reasons such as data missing and outliers, the size of the effective data sets varies between WFs. However, regardless of the size of the dataset, 70% is used as the training set and 30% as the test set in the ML algorithm. During the training stage, the '*k-fold*' algorithm (k=5) is used to optimize the ML hyperparameters [33].

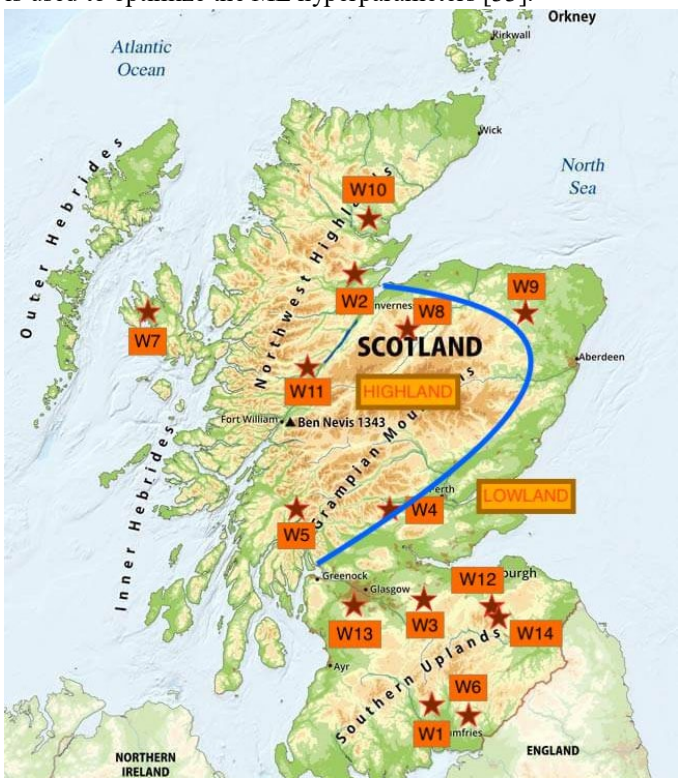

Figure 4. Wind farm locations

The wind data set comes from a long term hindcast performed with the Weather Resource and Forecast (WRF) model. This was applied to UK wind profiles over the period 2000-2010 [35] and provides hourly wind speed and direction for the entire UK region, at a spatial resolution of 3 km x 3 km. The hourly power generation data is derived from ELEXON settlement data [36] and pre-processed by Harrison [37], who transforms the half-hourly demand data to hourly resolution by taking hourly demand as the maximum of the two half-hour periods. Apart from these datasets, key parameters for WT at each WF, as well as WF latitude and longitude were collected. Table 1 and Figure 4 show the key information about these WFs. The blue line in Figure 4 indicates the boundary between the Highland and Lowland areas.

A. data pre-processing

Sometimes the power generation recordings may be empty or abnormal due to equipment maintenance, communication system damage, etc. Thus, before using the data to train the model the data set was cleaned up, which can be divided into two steps. First, all data records that contain negative values are deleted, simply since it is impossible to be lower than zero for whatever wind speed, power generation or wind direction. Secondly, the datasets outlier, which is defined as elements of more than three standard deviations σ (Eq. 13) from the mean μ (Eq.14), is removed [38].

$$\mu = \sum_{i=1}^{n} x_i / n \tag{13}$$

$$\sigma = \sqrt[2]{\sum_{i=1}^{n} (x_i - \mu)^2 / n} \tag{14}$$

where x_i are individual observations from a total n.

TABLE 1. WIND FARM INFORMATION

WF	Capacity (MW)	Number of Data Points	Cut in Speed (m/s)	Rated speed (m/s)	Cut out Speed (m/s)
W1	36.8	14692	3.5	13	25
W2	29.75	14558	3.5	16	25
W3	118	14079	3.5	13	25
W4	74	14335	3.5	14.5	25
W5	15	9969	3	13	25
W6	30	17015	3.5	12	22
W7	42	6325	2	15.5	25
W8	92	14942	3.5	17	25
W9	26.7	14647	2.5	18	25
W10	47.5	8036	3	13.5	25
W11	50	17059	3	13.5	25
W12	27.6	5221	3.5	13	25
W13	305	15601	3	12	20
W14	29.75	9704	3	14	25

B. Single wind farm modelling results

Tables 2 to 5 illustrate the statistical results of these models. It is clear that the SVR model is better than the basic SAPC model that relies on wind speed alone. Only four WFs have R^2 higher than 0.7 in SAPC, which are W1, W6, W7 and W9, while W4, W5 are lower than 0.6. However, with SVR, only the R^2 of W5 is lower than 0.6 and the ten WFs R^2 values are over 0.7. Generally speaking, the higher the R^2, the smaller differences between the observed data and the fitted values. From the perspective of RMSE, all SVR models' general results are lower than 0.2, with a minimum of 0.111 at W6. For SAPC, half of the WF RMSE is higher than 0.2, with a maximum of 0.307 at W12 and W6 has the lowest RMSE value of 0.138. Lower values of RMSE indicate better model fit.

Figure 5 shows the differences between the two models more intuitively. Except for W10, all the SAPC model results are 'above' the real generation and SVR results, which means the SAPC can overestimate the WF power generation. It is possible that SAPC does not use wind direction as an input to calibrate the model performance. In reality, for a given WF, the same background wind speed blowing from different directions is not equivalent: the influence of the topography, and differences in roughness, and wake effects affect the performance of each WT.

More evidence supports this in Tables 4 and 5, which break down the results into different directional sectors. Obviously, the results of RMSE and R^2 vary along with different wind directions in both SAPC and SVR models. For example, in W14, the RMSE value in direction 180°-225° in the SAPC model can reach up to 0.381, with an R^2 value down to 0.554, the worst performing case. In the SVR model, after introducing wind direction as an independent variable, RMSE and R^2 values still change, but the variation is less than that in the SAPC model. It shows that wind direction is a crucial factor in WF power generation modelling.

Similarly, Figure 6 indicates the directional difference in capacity factor among real generation, SVR model and SAPC model. The diagram is obtained by averaging the total power generation capacity factor/power curve in a specific direction range where the SAPC modelled output is regarded as one unit in every direction. From Figure 6, it is clear that the actual output is generally about 40%-80% of the SAPC output and the specific value varies with the direction of the wind and WF. The extreme cases are W11 and W8 where the actual output under different wind directions varies from 70% to 140% of the corresponding SAPC output.

Overall, the changing RMSE and R^2 means that the actual wind speed experienced by the WT is not the same as the hindcast value. This is entirely reasonable given that hindcasts themselves are not perfectly accurate, although the specific modelled data has been well validated against a wind variety of WF and met station anemometers. However, it is clear that in most cases the SVR captures the overall and directional features of the WF power output very well and shows up the limitations of simple power curves.

TABLE 2. SVR MODEL RMSE VALUE

RMSE SVR	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	w13	w14
Overall RMSE	0.126	0.172	0.118	0.152	0.189	0.111	0.146	0.160	0.138	0.173	0.140	0.155	0.130	0.130
angle (0:45)	0.131	0.123	0.082	0.147	0.225	0.118	0.147	0.201	0.149	0.123	0.154	0.125	0.084	0.130
angle (45:90)	0.115	0.124	0.113	0.152	0.224	0.116	0.094	0.111	0.098	0.121	0.101	0.170	0.142	0.104
angle (90:135)	0.142	0.118	0.188	0.119	0.276	0.113	0.147	0.150	0.086	0.203	0.101	0.099	0.136	0.142
angle (135:180)	0.116	0.176	0.107	0.079	0.169	0.116	0.151	0.193	0.165	0.205	0.167	0.141	0.112	0.107
angle (180:225)	0.154	0.238	0.140	0.181	0.137	0.108	0.196	0.175	0.161	0.149	0.162	0.184	0.132	0.146
angle (225:270)	0.115	0.170	0.117	0.160	0.159	0.111	0.190	0.189	0.162	0.178	0.148	0.186	0.166	0.142
angle (270:315)	0.113	0.173	0.098	0.124	0.172	0.102	0.128	0.123	0.126	0.153	0.112	0.183	0.109	0.135
angle (315:360)	0.089	0.137	0.065	0.177	0.157	0.110	0.109	0.098	0.107	0.186	0.130	0.120	0.082	0.117

TABLE 3. SAPC MODEL RMSE VALUE

RMSE SAPC	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	w13	w14
Overall RMSE	0.168	0.210	0.274	0.220	0.236	0.138	0.220	0.161	0.157	0.201	0.155	0.307	0.276	0.278
angle (0:45)	0.191	0.235	0.296	0.163	0.148	0.157	0.211	0.159	0.172	0.227	0.150	0.228	0.197	0.196
angle (45:90)	0.191	0.178	0.311	0.149	0.234	0.179	0.192	0.127	0.182	0.154	0.125	0.294	0.277	0.196
angle (90:135)	0.176	0.143	0.315	0.147	0.380	0.117	0.243	0.175	0.125	0.229	0.152	0.240	0.248	0.155
angle (135:180)	0.140	0.192	0.296	0.131	0.255	0.098	0.226	0.192	0.161	0.239	0.190	0.316	0.242	0.235
angle (180:225)	0.141	0.227	0.297	0.247	0.235	0.105	0.168	0.165	0.164	0.187	0.168	0.364	0.310	0.381
angle (225:270)	0.198	0.237	0.252	0.254	0.170	0.111	0.183	0.179	0.171	0.185	0.163	0.335	0.262	0.350
angle (270:315)	0.168	0.217	0.235	0.279	0.200	0.162	0.283	0.142	0.158	0.155	0.150	0.285	0.341	0.274
angle (315:360)	0.103	0.194	0.194	0.202	0.203	0.163	0.226	0.116	0.140	0.195	0.120	0.329	0.211	0.200

TABLE 4. SVR MODEL R^2 VALUE

R^2 SVR	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	w13	w14
Overall R^2	0.769	0.639	0.754	0.657	0.520	0.776	0.741	0.710	0.749	0.698	0.734	0.756	0.785	0.772
angle (0:45)	0.544	0.403	0.355	0.619	0.486	0.468	0.515	0.677	0.463	0.674	0.514	0.419	0.593	0.687
angle (45:90)	0.768	0.309	0.707	0.752	0.535	0.771	0.737	0.525	0.678	0.764	0.716	0.690	0.711	0.720
angle (90:135)	0.835	0.901	0.685	0.877	0.229	0.923	0.702	0.819	0.943	0.682	0.847	0.955	0.893	0.857
angle (135:180)	0.757	0.604	0.702	0.832	0.450	0.779	0.816	0.621	0.734	0.500	0.674	0.854	0.821	0.893
angle (180:225)	0.766	0.515	0.704	0.595	0.651	0.805	0.764	0.720	0.645	0.704	0.789	0.705	0.781	0.676
angle (225:270)	0.828	0.639	0.807	0.641	0.728	0.829	0.791	0.714	0.651	0.797	0.770	0.749	0.771	0.773
angle (270:315)	0.678	0.650	0.737	0.534	0.563	0.649	0.703	0.716	0.781	0.775	0.705	0.688	0.693	0.687
angle (315:360)	0.569	0.605	0.522	0.392	0.394	0.538	0.785	0.418	0.774	0.602	0.597	0.462	0.574	0.652

TABLE 5. SAPC MODEL R^2 VALUE

R^2 SAPC	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14
Overall R^2	0.737	0.626	0.644	0.553	0.506	0.730	0.730	0.699	0.738	0.680	0.662	0.612	0.665	0.674
angle (0:45)	0.511	0.459	0.307	0.617	0.599	0.422	0.565	0.702	0.496	0.653	0.491	0.388	0.550	0.632
angle (45:90)	0.740	0.335	0.620	0.777	0.547	0.758	0.735	0.536	0.699	0.744	0.708	0.579	0.621	0.668
angle (90:135)	0.852	0.898	0.582	0.858	0.214	0.930	0.691	0.822	0.932	0.690	0.786	0.837	0.802	0.887
angle (135:180)	0.735	0.623	0.542	0.821	0.415	0.794	0.794	0.624	0.745	0.482	0.566	0.708	0.708	0.805
angle (180:225)	0.773	0.531	0.629	0.563	0.650	0.793	0.784	0.729	0.633	0.701	0.715	0.585	0.656	0.554
angle (225:270)	0.788	0.642	0.761	0.610	0.732	0.812	0.783	0.716	0.657	0.784	0.728	0.603	0.704	0.649
angle (270:315)	0.681	0.654	0.657	0.470	0.579	0.622	0.647	0.715	0.756	0.781	0.708	0.595	0.530	0.639
angle (315:360)	0.540	0.588	0.500	0.246	0.290	0.557	0.745	0.395	0.746	0.598	0.581	0.416	0.558	0.648

TABLE 6. SVR CROSS TESTING RESULTS R^2 (COLORED VALUES MEAN BETTER THAN SAPC MODEL)

Model / Apply	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14
W1		0.626	0.663	0.527	0.435	0.738	0.748	0.675	0.725	0.669	0.708	0.592	0.749	0.599
W2	0.707		0.716	0.491	0.404	0.694	0.703	0.563	0.721	0.644	0.590	0.686	0.747	0.750
W3	0.664	0.576		0.425	0.338	0.571	0.724	0.553	0.711	0.584	0.550	0.734	0.736	0.695
W4	0.602	0.523	0.573		0.356	0.558	0.623	0.508	0.648	0.500	0.515	0.596	0.695	0.687
W5	0.698	0.623	0.617	0.407		0.680	0.591	0.610	0.645	0.727	0.628	0.468	0.684	0.448
W6	0.754	0.633	0.640	0.523	0.419		0.715	0.630	0.711	0.647	0.679	0.544	0.718	0.637
W7	0.734	0.627	0.692	0.572	0.420	0.690		0.625	0.749	0.644	0.647	0.683	0.773	0.706
W8	0.751	0.628	0.602	0.505	0.488	0.714	0.688		0.695	0.697	0.720	0.538	0.713	0.557
W9	0.720	0.630	0.739	0.490	0.382	0.683	0.755	0.574		0.639	0.596	0.700	0.761	0.732
W10	0.719	0.618	0.675	0.462	0.481	0.706	0.679	0.606	0.689		0.623	0.582	0.707	0.627
W11	0.762	0.602	0.605	0.481	0.463	0.713	0.715	0.710	0.661	0.672		0.526	0.717	0.498
W12	0.625	0.548	0.693	0.525	0.403	0.571	0.668	0.481	0.652	0.560	0.509		0.724	0.719
W13	0.717	0.617	0.723	0.551	0.374	0.662	0.749	0.586	0.724	0.601	0.613	0.708		0.712
W14	0.642	0.579	0.697	0.546	0.354	0.621	0.684	0.489	0.686	0.563	0.515	0.726	0.745	

TABLE 7. SVR CROSS TESTING RESULTS RMSE (COLORED VALUES MEAN BETTER THAN SAPC MODEL)

Model / Apply	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14
W1		0.207	0.196	0.208	0.132	0.132	0.189	0.151	0.183	0.148	0.231	0.146	0.225	0.126
W2	0.151		0.157	0.191	0.211	0.152	0.165	0.224	0.147	0.189	0.180	0.210	0.131	0.179
W3	0.200	0.219		0.208	0.243	0.211	0.161	0.270	0.177	0.242	0.218	0.162	0.155	0.155
W4	0.179	0.215	0.169		0.225	0.183	0.178	0.263	0.175	0.235	0.203	0.200	0.147	0.158
W5	0.161	0.182	0.155	0.199		0.160	0.182	0.224	0.169	0.193	0.183	0.226	0.147	0.203
W6	0.130	0.181	0.235	0.215	0.220		0.216	0.186	0.185	0.203	0.144	0.295	0.170	0.262
W7	0.169	0.196	0.134	0.172	0.216	0.178		0.249	0.155	0.212	0.196	0.177	0.135	0.147
W8	0.144	0.193	0.275	0.242	0.214	0.177	0.236		0.191	0.191	0.128	0.298	0.199	0.288
W9	0.159	0.177	0.141	0.192	0.220	0.169	0.146	0.236		0.195	0.188	0.189	0.128	0.172
W10	0.140	0.175	0.191	0.211	0.194	0.138	0.182	0.203	0.165		0.163	0.256	0.152	0.224
W11	0.127	0.178	0.219	0.209	0.198	0.133	0.178	0.175	0.165	0.181		0.236	0.159	0.242
W12	0.184	0.210	0.136	0.181	0.219	0.189	0.164	0.263	0.175	0.231	0.209		0.143	0.150
W13	0.170	0.199	0.129	0.177	0.224	0.182	0.142	0.251	0.160	0.219	0.199	0.167		0.148
W14	0.189	0.210	0.132	0.180	0.230	0.191	0.161	0.269	0.168	0.230	0.214	0.162	0.147	

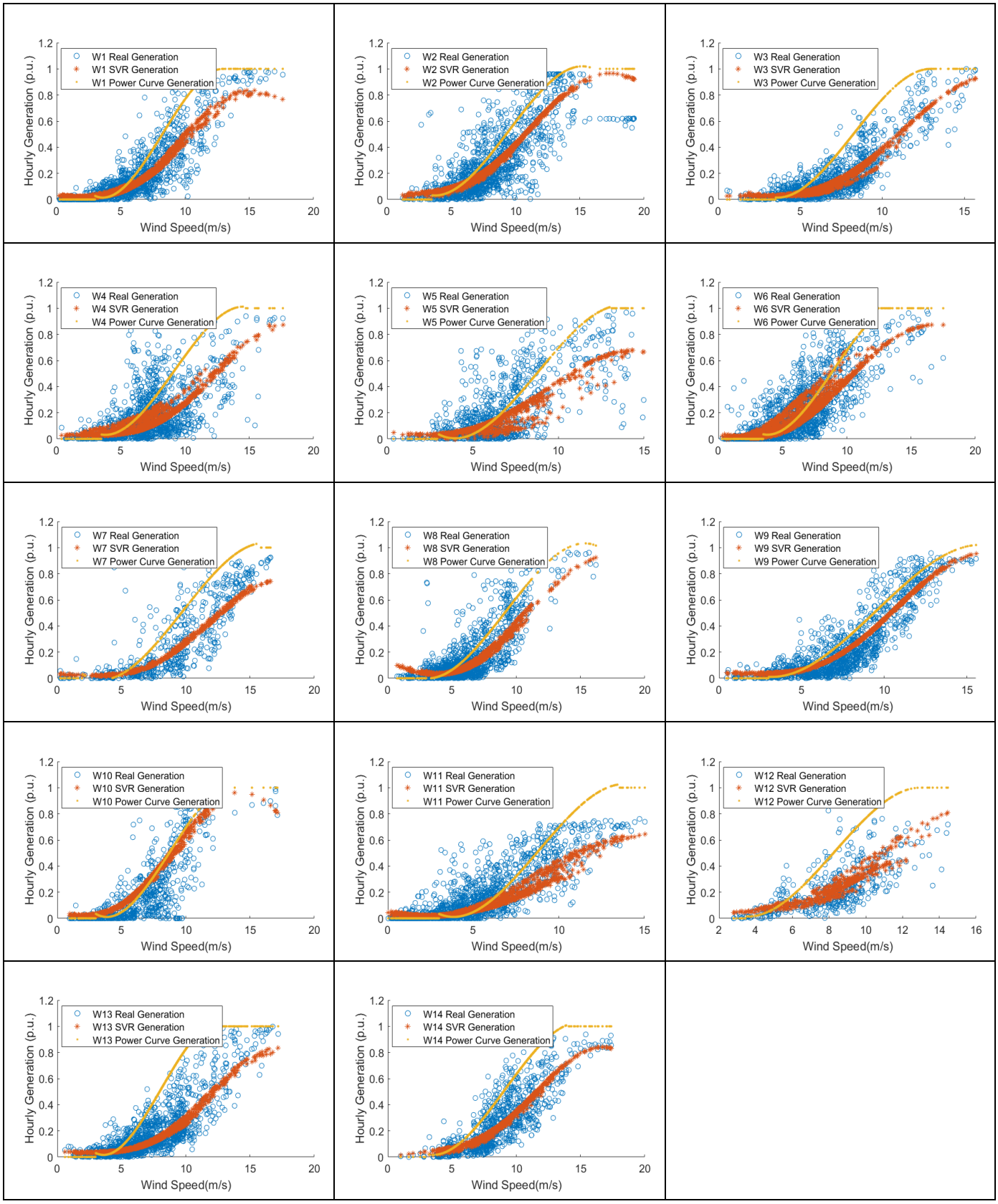

Figure 5. SVR and power curve modelling results in all 14 WFs

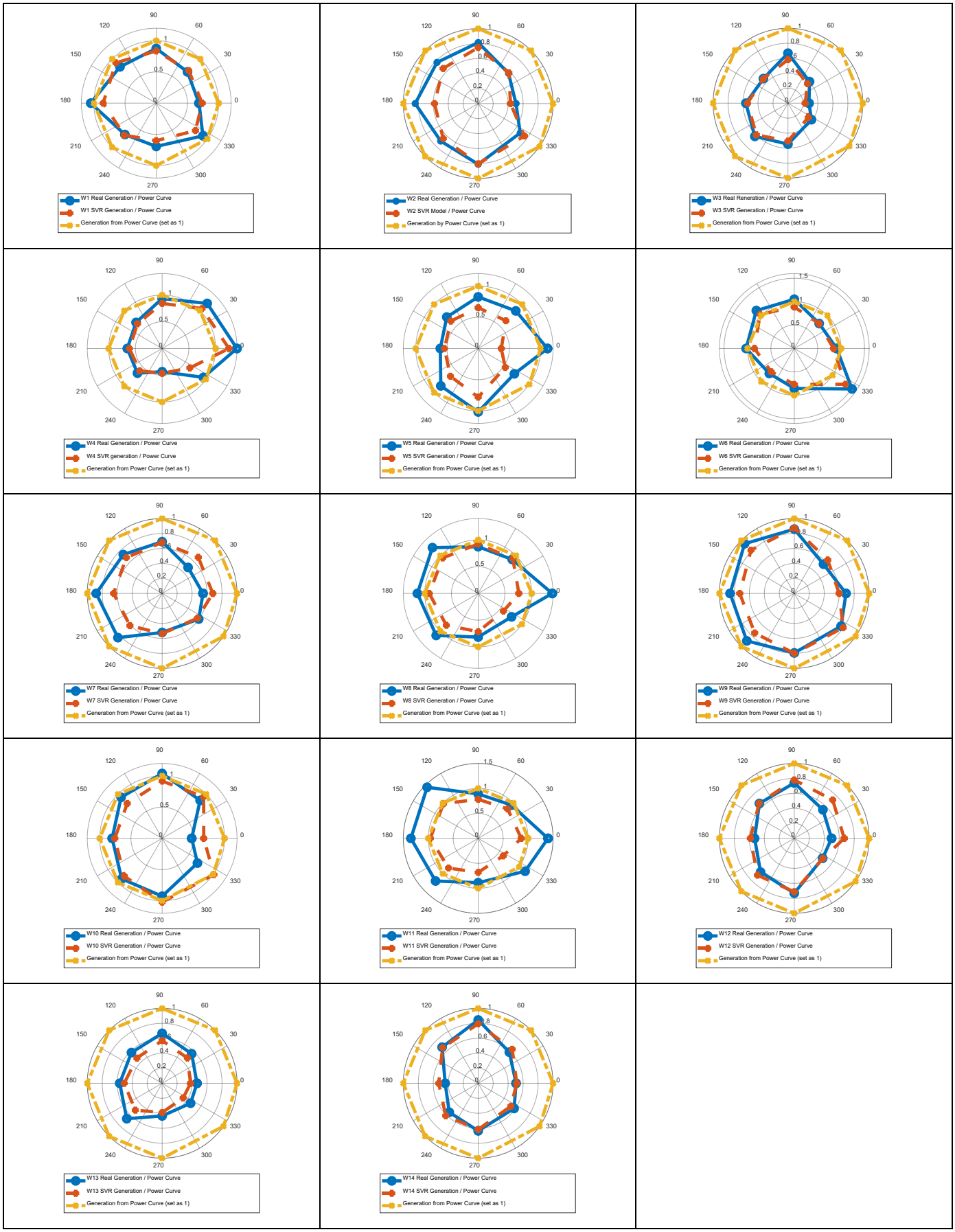

Figure 6. Modelling results difference by direction

C. Transfer modelling strategy

Until now, the planning stage WF power generation modelling can only rely on the power curve provided by the manufacturer. Here, it is anticipated that an existing WF's operating data can help to improve the accuracy of other planning stage WF power output estimation. The following content presents the facts and inference based on the application results of this strategy.

This section applies a novel model transfer strategy to WF modelling in the planning stage. A series of one-to-one crossover experiments are performed on all WF models, and the results are corrected using the ratio of power curves in two WFs. Table 6 and Table 7 demonstrate the main results. Here, a criterion is proposed to judge the success of this strategy:

> *The SVR model A of WF A and the SVR model B of WF B must obtain higher R^2 and RMSE values on both A and B than the SAPC model.*

This strategy is successfully applied to some WFs, which include W1-W6, W3-W13-W14-W12 and W7-W9. W1, W6, W12 and W14 belong to the southern uplands. W3 and W13 sit in the central lowland with almost the same latitudes. W9 is located in the eastern lowland and far from other WFs. W7, with relatively flat and low-lying terrain, is on the largest and northernmost island of the Inner Hebrides. Their positional relationship is shown in the following Table 8, which also shows two sets of WFs which marginally failed in the application of this strategy, W2-W10 and W8-W11. In order to illustrate the results more straight forwardly, Figure 7 graphically depicts one of the successful applications of the strategy.

TABLE 8. SUCCESSFUL WFS RELATIONSHIP BETWEEN EACH OTHER

Successful WF	Distance (km)	Latitude difference	Longitude difference
W1-W6	23.37	0.072	0.438
W7-W9	223.6	0.025	3.768
W3-W12	47.74	0.002	0.888
W3-W13	36.8	0.067	0.565
W3-W14	49.74	0.039	0.899
W12-W14	5.12	0.037	0.011
W12-W13	91.08	0.068	1.452
W13-W14	92.46	0.106	1.463
W8-W11 (marginally success)	45.68	0.191	0.738
W2-W10 (marginally success)	29.73	0.237	0.273

The distance between W12-W14 is only 5.12km, which is even less than the maximum distance between WTs in some large WFs. Therefore, they could arguably be regarded as one WF and it is not strange that the strategy can be successfully applied to them. Also, if any WF model, e.g. W3 and W13, transfers to one of them successfully, it should conceivably work on another one as well. In addition, W3 sits in the middle of central Scotland, its SVR model achieves more excellent results than the SAPC model in all WFs (Table 6 and Table 7). If one WF was chosen to represent the whole of Scotland's WFs, W3 would be the solid choice. However, except for W12, W13 and W14, the rest of the WF models cannot achieve a better result over the W3 SAPC model in modelling W3 power generation. These four WFs are in the Scottish Lowlands within relatively close proximity to W3. Moreover, W2-W10 and W8-W11 have similar conditions to all successful WFs above, but they are in the Scottish Highlands. Consequently, the strategy almost succeeds in these two cases: The R^2 value of the W8-W11 interactions is better than corresponding SAPC models, but there are still certain gaps in the RMSE value. In contrast, the W2-W10 interactions obtained superior RMSE values, but not R^2 results.

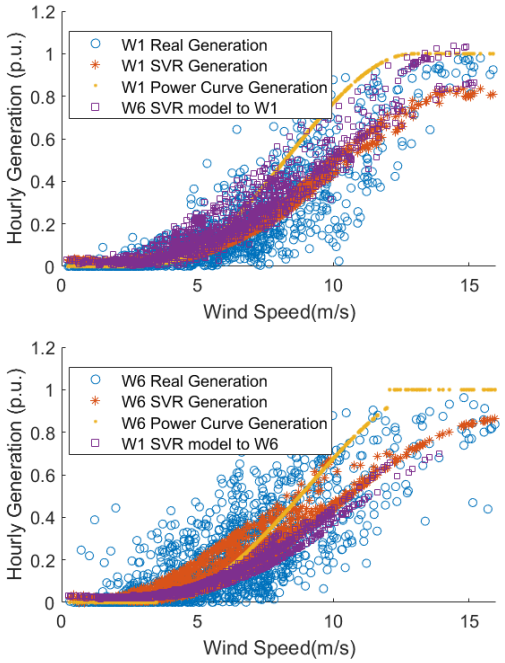

Figure 7. Example of successful cross testing model result

It is unexpected that W7-W9 is one of the successful pairings as the distance between them is the farthest among all successful combinations. Some characteristics may contribute to their models' excellent performance: W7 is located on an island in the Highlands, but like W9, its terrain is relatively flat. At the same time, the latitude difference between these two WFs is the second smallest of all successful WFs, only 0.025 degrees.

D. General facts from successful cases

After successful applications of the transfer strategy in these WFs, it is found that the successful cases meet the following necessary conditions. First, none of the successful cases contains a WF located in the Highlands. Therefore, it is probable that the terrain conditions have affected the success of the strategy. The mountainous area with complex terrain increases the specificity of the ML WF model and reduces the generalization ability of the model, which then fails to simulate the output of other WFs. Alongside this, the performance of the WRF hindcast model was seen to be poorer in very complex terrain than in flatlands [39].

Second, all these successful cases have and only have WFs within 0.2 degrees of latitude difference (Table 8). Probably, this is because the prevailing wind in Scotland is from the west to the south-west, hence WFs at similar latitudes enjoy similar wind regimes. But this could also actually mean that WFs with similar altitudes are more likely to apply this strategy successfully, since in Scotland, little latitude difference generally indicates small altitude gaps. Moreover, most WFs are within 50km of each other, but there are exceptions, such as W7-W9 (223km), W12-W13 (91km), and W13-W14 (92km).

Last but not least, it seems that longitude has little influence on the success of the model transfer strategy. The longitude difference between successful examples ranges from 0.1 to 1.5 degrees, which almost equals the width of Scotland. With the west to east direction of weather systems, longitude would tend to represent a timing difference between WF; it is probable that this is captured well by the SVR.

Overall, the terrain is the primary factor influencing whether the strategy can be successfully transferred between WFs or not, followed by latitude and distance. Longitude may not influence the application of the strategy. But due to too few examples (only 14 WFs), further work is needed. Based on the above findings, the most important inference of this section is the criteria indicating whether the WF can adopt the transfer strategy proposed in this work. First of all, both operating and planning stage WFs are in lowland areas. Secondly, within 50km of the WF there is an already operating WF with actual output data. Thirdly, the latitude difference between them is modest. On the premise of satisfying the above conditions, the target WF power output can be obtained by modelling the existing WF and applying this model transfer strategy.

IV. Conclusion

Numerous studies have been able to simulate the output of WFs. Obviously, WF modelling in the planning stage ("data blindness") still depends on the power curve and lacks the use of local WF information. In order to improve the accuracy of WF modelling in the planning stage, this paper proposes a novel method to directly migrate similar WF models to the planning WF with the help of the power curve. The particular emphasis is on a method that enables the transfer of information for existing WFs to potential sites where no measured data exists. By modelling 14 WFs distributed in Scotland, there is indeed a large gap between the WF output theoretically and the actual power generation. The lack of wind direction information within the power curve is a major reason for this phenomenon. Consequently, those models only based on wind speed result in a larger overall error. This work establishes an SVR model for WF and achieves better performance over the SAPC model from all perspectives.

Moreover, this study also discusses the possibility of sharing one data-driven ML model among different prospective WFs. For two WFs that are not in a mountainous area, they can share one ML model on the condition that the latitude difference between them is less than 0.2 degrees and the distance between them is less than 50km. After adjusting the results by the ratio of the power curve in these two WFs, the results of the ML model remain superior to the results of the SAPC. This suggests that if a WF is built near an existing WF, the power output of the target WF can be estimated using the SVR model of the existing WF, rather than solely relying on the manufacturer's power curves. Such a model could help WF investment decision-makers to estimate WF output more accurately at the planning stage, facilitating the successful final deployment. Future work will be on extending this developed method to other countries and observing the universality and effectiveness of the strategy.

Reference

[1] Walker, P., Mason, R., & Carrington, D. (2019). Theresa May commits to net zero UK carbon emissions by 2050. *The Guardian*, *11*(6), 19.

[2] BEIS, Department for Business, Energy & Industrial Strategy. "Energy consumption in the UK (2020)"

[3] Sun, W., & Harrison, G. P. (2019). Wind-solar complementarity and effective use of distribution network capacity. *Applied Energy*, *247*, 89-101.

[4] BEIS, Department for Business, Energy & Industrial Strategy. (2018). MODELLING 2050: ELECTRICITY SYSTEM ANALYSIS.

[5] Wang, X., Guo, P., & Huang, X. (2011). A review of wind power forecasting models. Energy procedia, 12, 770-778.

[6] Vargas, S. A., Esteves, G. R. T., Maçaira, P. M., Bastos, B. Q., Oliveira, F. L. C., & Souza, R. C. (2019). Wind power generation: a review and a research agenda. journal of cleaner production, 218, 850-870.

[7] Wan, Y. H., Ela, E., & Orwig, K. (2010). Development of an equivalent wind plant power-curve (No. NREL/CP-550-48146). National RenewaBLE ENERGY LAB.(NREL), GOLDEN, CO (UNITED STATES).

[8] Hayes, B. P., Ilie, I. S., Porpodas, A., Djokic, S. Z., & Chicco, G. (2011, June). Equivalent power curve model of a wind farm based on field measurement data. In *2011 IEEE Trondheim PowerTech* (pp. 1-7). IEEE.

[9] Wharton, S., & Lundquist, J. K. (2012). Assessing atmospheric stability and its impacts on rotor‐disk wind characteristics at an onshore wind farm. *Wind Energy*, *15*(4), 525-546.

[10] Jin, T., & Tian, Z. (2010, June). Uncertainty analysis for wind energy production with dynamic power curves. In *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems* (pp. 745-750). IEEE.

[11] Foley, A. M., Leahy, P. G., Marvuglia, A., & McKeogh, E. J. (2012). Current methods and advances in forecasting of wind power generation. *Renewable Energy*, *37*(1), 1-8.

[12] Deng, Y. C., Tang, X. H., Zhou, Z. Y., Yang, Y., & Niu, F. (2021). Application of machine learning algorithms in wind power: a review. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 1-22.

[13] Marvuglia, A., & Messineo, A. (2012). Monitoring of wind farms' power curves using machine learning techniques. *Applied Energy*, *98*, 574-583.

[14] He, D., & Liu, R. (2012, June). Ultra-short-term wind power prediction using ANN ensemble based on PCA. In *Proceedings of The 7th International Power Electronics and Motion Control Conference* (Vol. 3, pp. 2108-2112). IEEE.

[15] Yin, X., & Zhao, X. (2019). Big data driven multi-objective predictions for offshore wind farm based on machine learning algorithms. *Energy*, *186*, 115704.

[16] Simao, A., Densham, P. J., & Haklay, M. M. (2009). Web-based GIS for collaborative planning and public participation: An application to the strategic planning of wind farm sites. Journal of environmental management, 90(6), 2027-2040.

[17] Lledó, L., Torralba, V., Soret, A., Ramon, J., & Doblas-Reyes, F. J. (2019). Seasonal forecasts of wind power generation. *Renewable Energy*, *143*, 91-100.

[18] M.C. Alexiadis, P.S. Dokopoulos, H.S. Sahsamanoglou. Wind speed and power forecasting based on spatial correlation models IEEE Trans. Energy Convers., 14 (1999), pp. 836-842

[19] Ren, G., Wan, J., Liu, J., & Yu, D. (2020). Spatial and temporal correlation analysis of wind power between different provinces in China. *Energy*, *191*, 116514.

[20] Li, F., Ren, G., & Lee, J. (2019). Multi-step wind speed prediction based on turbulence intensity and hybrid deep neural networks. *Energy Conversion and Management*, *186*, 306-322..
[21] Ezzat, A. A., Jun, M., & Ding, Y. (2018). Spatio-temporal asymmetry of local wind fields and its impact on short-term wind forecasting. *IEEE transactions on sustainable energy*, *9*(3), 1437-1447..
[22] Sahin, A. D., & Sen, Z. (2000). Wind energy directional spatial correlation functions and application for prediction. *Wind Engineering*, *24*(3), 223-231.
[23] Damousis, I. G., Alexiadis, M. C., Theocharis, J. B., & Dokopoulos, P. S. (2004). A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Transactions on Energy Conversion*, *19*(2), 352-361.
[24] Czisch, G., & Ernst, B. (2001). High wind power penetration by the systematic use of smoothing effects within huge catchment areas shown in a European example. *Windpower 2001*.
[25] Papaefthymiou, G., & Pinson, P. (2008, May). Modeling of spatial dependence in wind power forecast uncertainty. In *Proceedings of the 10th International Conference on Probablistic Methods Applied to Power Systems* (pp. 1-9). IEEE.
[26] Ali, M., Ilie, I. S., Milanović, J. V., & Chicco, G. (2010, July). Probabilistic clustering of wind generators. In *IEEE PES General Meeting* (pp. 1-6). IEEE.
[27] Girard, R., & Allard, D. (2013). Spatio‐temporal propagation of wind power prediction errors. *Wind Energy*, *16*(7), 999-1012.
[28] Siebert, N., & Kariniotakis, G. (2006, February). Reference wind farm selection for regional wind power prediction models. In *European Wind energy conference, EWEC 2006* (pp. 10-pages).
[29] Ishihara, T., Yamaguchi, A., Ogawa, T., Sakai, K., & Fujino, Y. (2007). An upscaling approach for the regional wind power forecasting. *Ratio*, *243*, 100-0.
[30] Lange, M. (2005). On the uncertainty of wind power predictions—Analysis of the forecast accuracy and statistical distribution of errors. *J. Sol. Energy Eng.*, *127*(2), 177-184.
[31] Noble, W. S. (2006). What is a support vector machine?. *Nature biotechnology*, *24*(12), 1565-1567.
[32] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199-222.
[33] Mathworks. (2019). Fit a Support Vector Machine Regression Model—MATLAB Fitr svm-MathWorks United Kingdom. Available online: https://uk.mathworks.com/help/stats/fitrsvm.html (accessed on 13 August 2019).
[34] Standard, I. E. C. (2017). 61400-12-1, 2017. Wind turbines—part, 12-1.
[35] Hawkins, S., & Harrison, G. (2010). A reanalysis of UK wind speeds using the WRF mesoscale model. In European Wind Energy Conference and Exhibition (EWEC).
[36] Harrison, G. P., Hawkins, S. L., Eager, D., & Cradden, L. C. (2015). Capacity value of offshore wind in Great Britain. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 229(5), 360-372.
[37] Harrison, G. (2020). Prof Gareth Harrison | School of Engineering. Retrieved 20 October 2020, online available from <https://www.eng.ed.ac.uk/about/people/prof-gareth-harrison>[Accessed 25 September 2020].
[38] Li, Z., Friedrich, D., & Harrison, G. P. (2020). Demand forecasting for a mixed-use building using agent-schedule information with a data-driven model. *Energies*, *13*(4), 780.
[39] Hawkins, S. L. (2012). High resolution reanalysis of wind speeds over the British Isles for wind energy integration. *Ph. D. Thesis*.

Zihao Li. He is currently the PhD student in the Institution of Energy System (IES), the university of Edinburgh.

Zihao Li's research interest is in the low carbon energy system planning and operation model development. He was worked as carbon analyst in the sustainability office of the university and contributed some publications for National Centre for Energy Systems Integration (CESI).

Wei Sun (S'11, M'15 IEEE) received the Ph.D. degree in power system engineering from the University of Edinburgh, U.K., where he is currently a Chancellor's Fellow on multi-vector energy systems integration.

Dr. Sun's research interest is in the development and application of optimization methods to the planning, design and operation of smart energy systems. He is currently working on building a multi-scale energy system integration architecture for National Centre for Energy Systems Integration (CESI) and previously as a lead researcher on Hydrogen's Value in the Energy system (HYVE). He has also contributed to EPSRC funded projects including Realising Energy Storage Technologies in Low-carbon Energy Systems (RESTLESS) and Adaptation and Resilience In Energy Systems (ARIES).

Yue Xiang (Senior Member, IEEE) received the B.S. degree in electrical engineering and its automation and the Ph.D. degree in power system and its automation from Sichuan University, Chengdu, China, in 2010 and 2016, respectively.
He was a Joint Ph.D. Student with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA, from 2013 to 2014. He was a Visiting Researcher with the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K., from 2019 to 2020. He is currently an Associate Professor with the College of Electrical Engineering, Sichuan University, Sichuan, China. His research interests include distribution network planning and optimal operation, electric vehicle integration, integrated energy system planning, and smart grids.

Gareth P. Harrison (M'02, SM'14 IEEE, Fellow IET) is Bert Whittington Chair of Electrical Power Engineering and Deputy Head of the School of Engineering at the University of Edinburgh. He holds a Bachelor's degree and a Doctorate from the same institution.

Professor Harrison leads research activity across a wide area including integration of renewable energy within multi-vector energy systems, renewable resource assessment, climate change impacts on energy systems; and carbon footprints of energy systems. He is Associate Director of the National Centre for Energy Systems Integration (CESI), was previously Principal Investigator of the Adaptation and Resilience in Energy Systems project and is currently a Co-investigator on a range of EPSRC and EU projects covering energy storage, hydrogen, conventional generation, and offshore renewable energy.