The Accuracy of Wind Energy Forecasts in Great Britain and the Prospects for Improvement

Kevin F. Forbes
Economic and Social Institute (ESRI) and Catholic University of America

Ernest M. Zampelli
The Catholic University of America
Washington, DC USA

Abstract

This paper assesses and offers a methodology for improving the accuracy of wind energy forecasts in Great Britain. We present evidence that the day-ahead wind energy forecast in Great Britain is substantially less accurate than generally believed. Evidence is also presented that the notifications of expected generation submitted by the wind energy generators to the system operator are inaccurate relative to the notifications provided by the operators of generating stations fuelled by conventional energy.

Regarding prospects for improving forecast accuracy, we present evidence that the wind energy forecasts do not fully reflect forecasted meteorological conditions. With this insight, we developed a revised wind energy forecasting model using time-series methods and forecasted meteorological conditions as regressors. The out-of-sample results are highly encouraging. Specifically, the out-of-sample weighted-mean-absolute-percentage error (WMAPE) of the period-ahead revised forecast is less than one-fifth the WMAPE of the existing near real-time projection.

Keywords
Wind Energy Forecasting, Forecast Accuracy

JEL Codes: Q42, Q47

Acronyms: ARMAX, autoregressive–moving-average with exogenous inputs; CCGT, combined cycle gas turbine; CF, capacity factor; CWRMSE, capacity weighted root-mean squared error, EWRMSE, energy weighted root mean squared error; FPN, Final Physical Notification; MAPE, mean average percentage error; MFP, multivariable fractional polynomial; NIV, net imbalance volume; WMAPE, weighted mean average percentage error;

1 Introduction

A growing literature has affirmed with great confidence the feasibility of high penetration levels of wind energy into the power grid under existing operational methods. This optimism is based in part on the belief that improvements in wind energy forecasting have effectively "solved" the challenge of wind energy's intermittency. For example, the UK's Royal Academy of Engineering has explicitly indicated that wind energy's capacity-weighted forecast error of about 5% is evidence that wind energy forecasts in Great Britain are highly accurate (Royal Academy of Engineering, 2014, p. 33*)*. Using data from Great Britain, we assess and challenge this claim. We then offer a methodology to improve short-run forecast accuracy.

The paper is organized as follows. In section 2, the implications of inaccurate forecasts for electric grid stability are discussed. Section 3 reviews and critiques the literature on the accuracy of wind energy forecasts while Section 4 compares the accuracy of day-ahead load and wind energy forecasts in Great Britain using the weighted-mean-absolute-percentage-error (WMAPE) measure of forecast accuracy due to its ease of interpretability and its independence in terms of scale. Using this metric, we find that the day-ahead wind energy forecast in Great Britain is substantially less accurate than the day-ahead load forecast.

In section 5 the accuracy of the final physical notifications of expected generation submitted by the wind energy generators to the system operator one hour prior to real time are compared to the accuracy of the final physical notifications submitted by generating stations fuelled by conventional energy. The final physical notifications submitted by the operators of the wind energy generating stations are shown to be substantially less accurate than the notifications supplied by the stations fuelled by conventional energy sources.

Section 6 considers, and answers in the affirmative, the question of whether an improvement in the accuracy of the wind energy forecasts is possible. Evidence is presented that the wind energy forecasts do not fully reflect the information contained in the day-ahead weather forecasts. The wind energy forecast errors also have the undesirable property of being correlated with the forecasted level of wind energy.

Section 7 presents a time-series econometric model of wind energy production. The model is estimated using half-hour data over the period 1 January 2012 - 31 December 2013. We conclude that a substantial improvement in near real-time forecast accuracy is possible by modifying the forecast information based on the time-series model. The findings are summarized in section 8.

2 The Importance of Accurate Forecasts

"Keeping the lights on" requires adherence to fairly stringent stability conditions in terms of system frequency, i.e., the level of voltage and current oscillations each second. System frequency falls (rises) when demand exceeds (or is less than) supply. System frequency in Great Britain varies around the target of 50 Hz. with National Grid, the system operator in Great Britain, is obligated to keep system frequency within 1% of the 50 Hz target, i.e. +/- 50 mHz. (United Kingdom Electricity Safety, Quality and Continuity Regulations, 2002). Deviations within the band +/- 20 mHz are considered normal. Violations of the operational limits +/- 20 mHz do occur (Figure 1). These violations pose a risk to equipment and the overall reliability of the electric power system. When a large and sustained violation occurs, generators and other equipment connected to the transmission system may be automatically disconnected to prevent equipment damage which in turn could result in an electricity blackout. An example of this risk was when a transmission line "tripped,", i.e. automatically disconnected itself from the transmission system to prevent damage, at about 17:15 local time on 9 November 1965 in the province of Ontario Canada. This adversely affected system frequency in New York State which led to generators going offline which in turn induced further declines in system frequency. By about 17:28 local time, system frequency was 11 Hz below its target (Federal Power Commssion, 1967, p. 49). The electric power system in the northeast United States and Canada was largely unable to cope and by 17:30 local time about 30 million individuals lost power (Federal Power Commission, 1967, p. 17). Given these possible consequences, it is troubling to note that the one-second system frequency data from National Grid (https://www.nationalgrideso.com/balancing-services/frequency-response-services/historic-frequency-data) indicates that the incidence of violations of the operational limits in 2018 was about 2.5 times the level in 2014 (36,325 seconds in 2018 vs 14,200 seconds in 2014). Fortunately, the statutory limits were not violated over the period 2014-2018 but the range of 49.556 to 50.41 Hz is hardly reassuring.

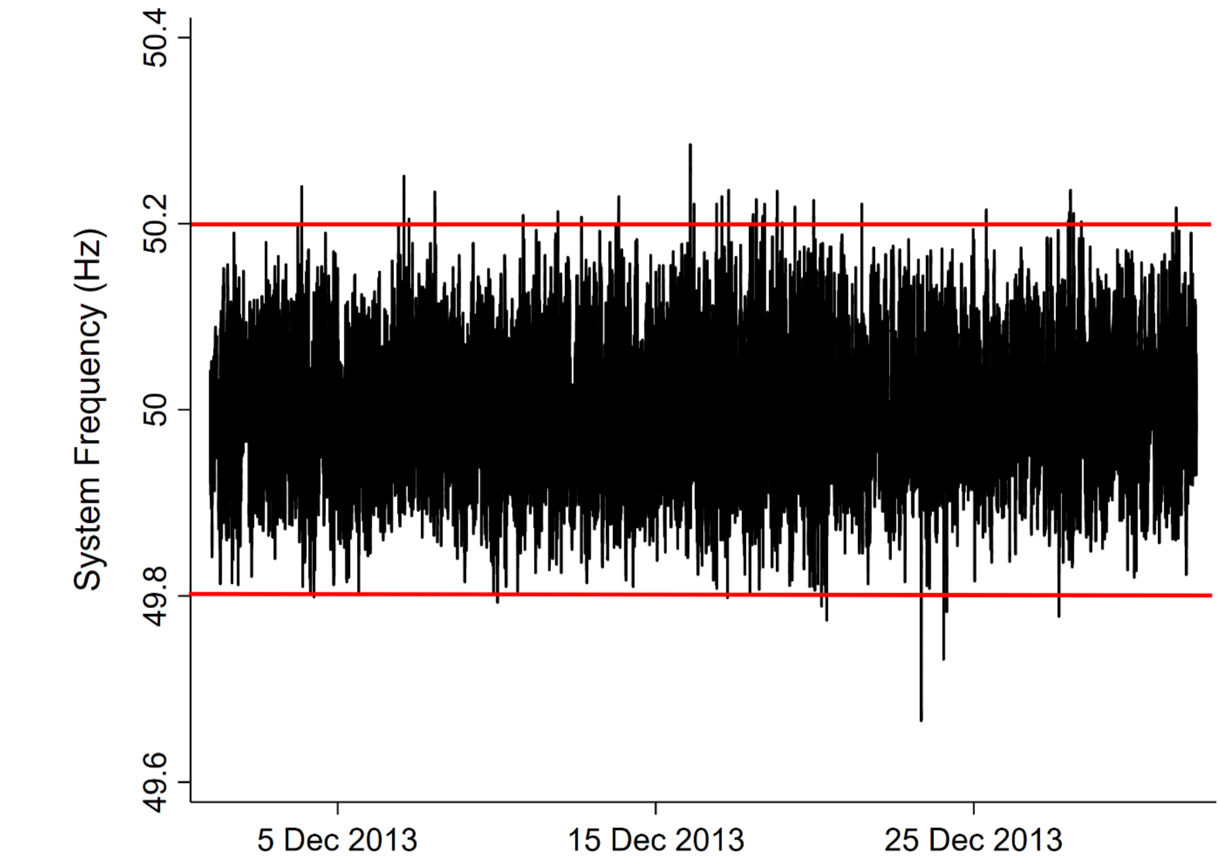

Figure 1: **System frequency in Great Britain's power grid, 1 Dec–31 Dec 2013**

Errors in electricity load forecasts and the failure of suppliers to adhere to their generation and transmission schedules are the root causes of the variability in system frequency. National Grid attempts to offset this by dispatching balancing power, supplied largely by combined-cycle gas turbines (CCGT) and coal-fired power stations. A key measure of the electricity imbalance during a 30-minute settlement period in Great Britain is known as the net imbalance volume (NIV). NIV equals the sum of all energy deployments initiated by National Grid to balance the system during the settlement period. Positive (negative) NIV values indicate a shortage (surplus). Large NIV outcomes in absolute value represent instances in which the reliability of the system is challenged. Large positive NIV values are especially problematic as most conventional generating technologies require non-negligible response times when additional generation is needed. Over the period 1 January 2012 through 31 December 2013, the average level of NIV was negative, but very volatile from one market period to the next (Figure 2).

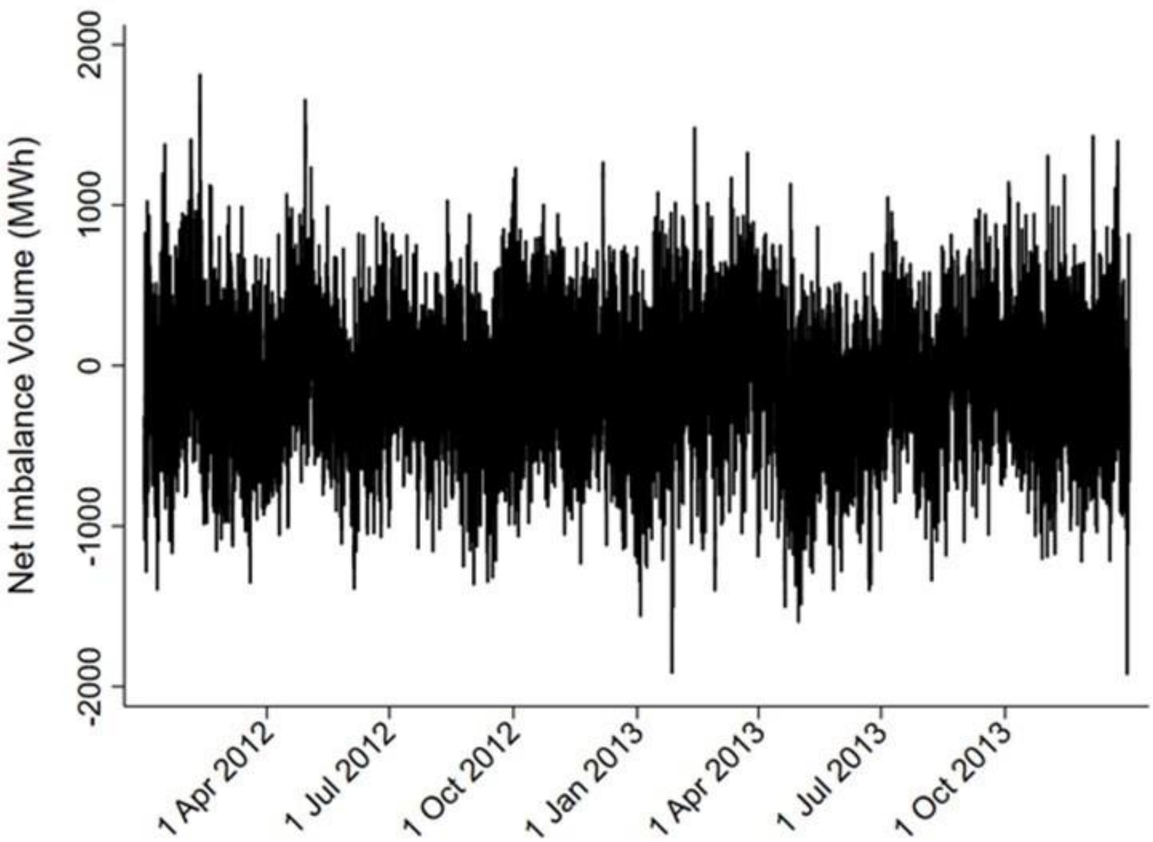

Figure 2: **Net energy imbalance in the power grid that serves Great Britain, 1 Jan 2012 - 31 Dec 2013**

The costs of balancing the system are nontrivial. Based on data downloaded from National Grid, the total balancing costs in 2012 and 2013 were about £923 million and £989 million, respectively (https://www.nationalgrid.com/uk/electricity/charging-and-methodology/balancing-services-use-system-bsuos-charges). This is over 50% higher than the annual costs in 2005-2006 (National Audit Office, 2014 p. 18). This finding of increasing costs is consistent with the results reported by Joos and Staffell (2018) for both Germany and Great Britain.

3 The Literature on Wind Energy Forecast Accuracy

As noted earlier, the UK's Royal Academy of Engineering (2014, p. 33) has reported a capacity weighted forecast error of about 5%. This is presented as evidence that wind energy forecasts are highly accurate. But is this really true? Before answering, consider that Lange, et al. (2006, 2007, 2009), Cali et al. (2006), Krauss, et al. (2006), Holttinen, et al. (2006), Kariniotakis, et al. (2006), Milligan, et al. (2009), IPCC (2012 p, 623) and even North American Electric Reliability Corporation (NERC; 2010, p. 9) all report wind energy forecast error metrics that are weighted by the capacity of the equipment used to produce the energy. These publications make it appear that wind energy forecasting has resolved the intermittency challenge. For example, in a publication entitled, "Wind Power Myths Debunked," Milligan, et al. (2009) draw on research from Germany to assert that it is not difficult to accurately forecast wind energy. In their words: "…typical wind forecast errors for a single wind project are 10% to 15% root mean-squared error (RMSE) **of installed wind capacity** (emphasis added) but drop to 5% to 7% for all of Germany."(Milligan, et al. 2009, p. 93).

The notion that the challenge of wind energy's intermittency is mythological is increasingly accepted. For example, Sovacool (2009) interviewed 62 individuals affiliated with electric utilities, regulatory agencies, interest groups, energy systems manufacturers, nonprofit organizations, energy consulting firms, universities, national laboratories, and state institutions in the United States. Based on their responses and a review of the literature, Sovacool concluded that "…the intermittency of renewables can be predicted, managed, and mitigated, and that the current technical barriers are mainly due to the social, political, and practical inertia of the traditional electricity generation system."

In light of these beliefs, we proceed by taking a closer look at the capacity weighted root mean squared error (CWRMSE), which can be calculated as follows:

$$CWRMSE = \frac{\sqrt{\frac{\sum_{t=1}^{T}(Actual_t - Forecast_t)^2}{T}}}{Installed\ Capacity} \quad *100\% \qquad (1)$$

Where $Actual_t$ is the actual wind energy produced in megawatts (MW) for period t, $Forecast_t$ is the period t forecasted level of energy in period t measured in MW of energy, and Installed Capacity is the MW of wind energy that could be produced if the wind turbines were operating at their capacity.

Unfortunately, the actual wind energy generated is almost always substantially less than the installed capacity of the wind turbines. On average, the median ratio of actual wind energy production to capacity, i.e., the capacity factor, in Great Britain was about 26% for the period 1 January 2012 through 31 December 2013. The highest capacity factor over the same period was about 92%. Thus, advocates for normalizing wind energy RMSEs by installed capacity are implicitly in favor of weighting by a benchmark that is much higher than the typical value of the actual level of generation. Moreover, meaningful comparisons between CWRMSEs and published measures of load forecast accuracy are precluded as the latter are normalized by the mean of actual energy consumption. An alternative metric of forecast accuracy is the energy weighted RMSE (EWRMSE) calculated as:

$$EWRMSE = \frac{\sqrt{\frac{\sum_{t=1}^{T}(Actual_t - Forecast_t)^2}{T}}}{(\sum_{t=1}^{T} Actual_t)/T} \quad *100\% \qquad (2)$$

where the denominator is average energy produced from t = 1 to T. Defining the average capacity factor (CF) as

$$CF = \frac{(\sum_{t=1}^{T} Actual_t)/T}{Installed\ Capacity}, \tag{3}$$

equation (2) can be rewritten to yield

$$\text{CWRMSE} = \text{CF} * \text{EWRMSE}. \tag{4}$$

The relationship described by (4) demonstrates that EWRMSE will exceed the associated CWRMSE whenever CF is less than one. The differences for wind energy can be substantial. For example, Staffell and Pfenninger (2016) estimate an average annual capacity factor for Great Britain of about 29% over the period 2005-2014. From (4), this means that the EWRMSE is almost 3.5 times larger than the associated CWRMSE. In effect, equation (4) shows explicitly that the CWRMSE conflates the forecast error normalized by average energy produced with the capacity factor. Consequently, a generating technology with a low capacity factor may have a smaller CWRMSE than a technology with higher capacity factor but a lower EWRMSE. Moreover, a decline in the capacity factor with EWRMSE unchanged would create the mistaken impression of improved forecast accuracy even though the forecast error in terms of energy has remained the same.

In this paper, we will not engage in a debate of EWRMSE vs. CWRMSE – we have discovered that adherents of capacity weighting simply justify this practice because it is the approach favored by the wind power industry – as if the important issue of wind energy integration should be under the preview of commercial interests. Instead, we will examine the claim of wind energy forecast accuracy using a modified accuracy metric that is well known and largely accepted by professional forecasters outside the field of wind energy forecasting. The metric we will employ is a variant of the mean-absolute-percentage-error (MAPE), which is calculated as follows:

$$MAPE = \sum_{t=1}^{T} \left| \frac{Forecast_t - Actual_t}{Actual_t} \right| *100\% \quad (5)$$

According to Hyndman (2006), a prominent scholar in the field of forecasting, MAPE is the most commonly used measure of forecast accuracy when one wants to express the error as a percent so as to compare forecast performance across different data sets. One shortcoming of MAPE is that the error for time period t is undefined if $Actual_t$ equals zero. Other shortcomings include a bias noted by Armstrong (1985) as well as the fact that the measure is sensitive to small errors that are large in percentage terms. To avoid these issues, we will make use of a weighted-mean-absolute-percentage-error (WMAPE) which is defined as follows:

$$WMAPE = \frac{\frac{1}{T}\sum_{t=1}^{T} |Actual_t - Forecast_t|}{\frac{1}{T}\sum_{t=1}^{T} Actual_t} *100\% \quad (6)$$

WMAPE represents the mean absolute error divided by the mean of the actual outcome (Kolassa and Schütz, 2007). In contrast to either the CWRMSE or the EWREMSE, it is relatively easily understood by users. It has the advantage of being scale-free and thus it facilitates comparison of forecast accuracy among forecast variables of interest such as solar energy, wind energy, and load.

4 The Accuracy of the Day-Ahead Load and Wind Energy Forecasts in Great Britain

The day-ahead wind energy forecast in Great Britain is substantially less accurate than the day-ahead load forecast. Specifically, the WMAPE for the day-ahead wind energy forecast is 31.43% (Figure 3), while the WMAPE corresponding to the day-ahead load forecast is 1.42% (Figure 4). The differential in the two levels of accuracy is even greater when the comparison is based on the MAPEs as opposed to the WMAPEs. Specifically, the MAPE corresponding to the day-ahead load forecast is 1.46% while the MAPE for the day-ahead wind energy forecast is about 58.66%. To double-check ourselves, we calculated a mean-squared-error skill score (MSESS) for both sets of forecast errors using the respective persistence forecasts as the reference, the latter being defined as a forecast that assumes the activity level in period t will equal the activity level in period t-1. The results indicate that the day-ahead load forecast is superior to its persistence load forecast, while the day-ahead wind energy forecast is inferior to a persistence wind energy forecast. This finding is consistent with the conclusions by Forbes et al. (2012) for Western Denmark, Eastern Denmark, Elia in Belgium, 50Hertz formerly Vattenfall) in Germany, Amprion (formerly RWE) in Germany, TenneT in Germany (formerly E.ON NETZ) the Midwest ISO in the U.S., and ERCOT in the U.S. For each of these electric power systems, the day-ahead load forecasts were more accurate than the day-ahead wind energy forecasts by a wide margin.

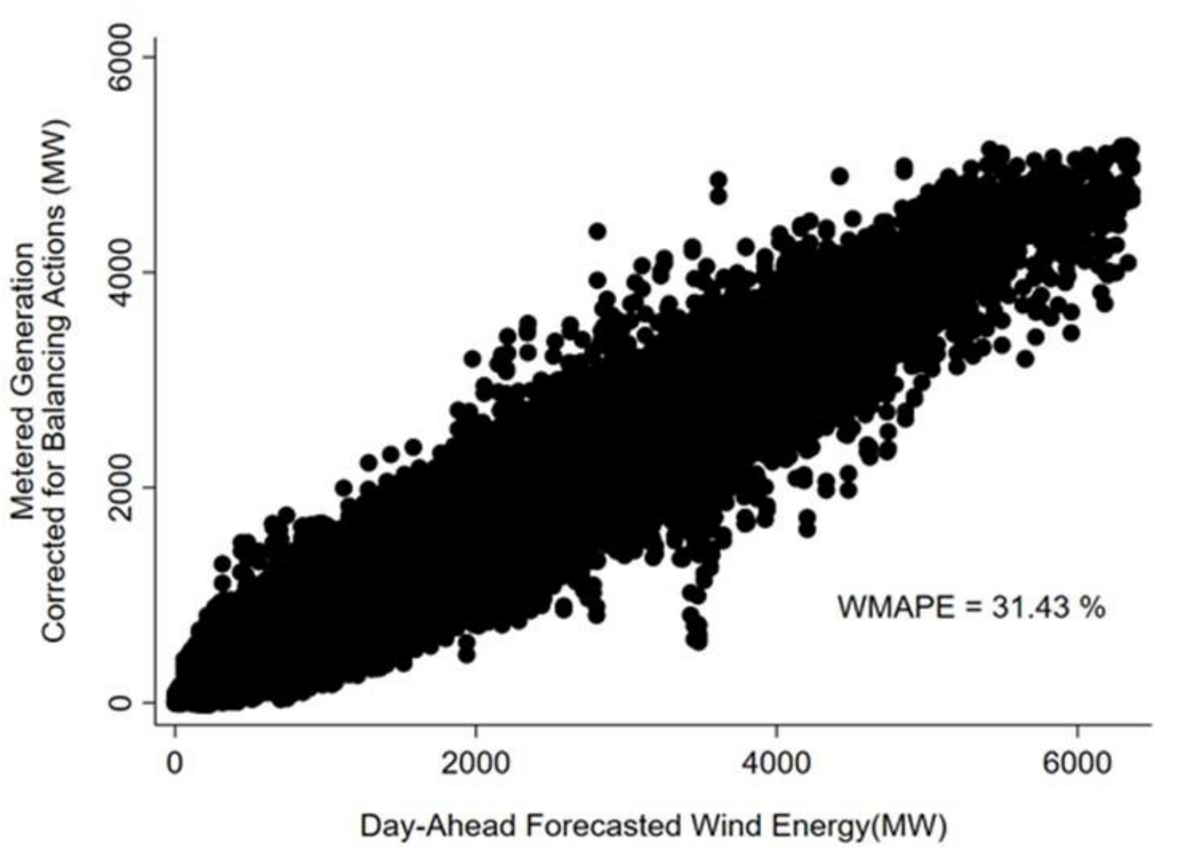

Figure 3: National Grid's day-ahead wind energy forecast and the metered level of wind energy generation adjusted for curtailments, 1 Jan 2012–31 Dec 2013

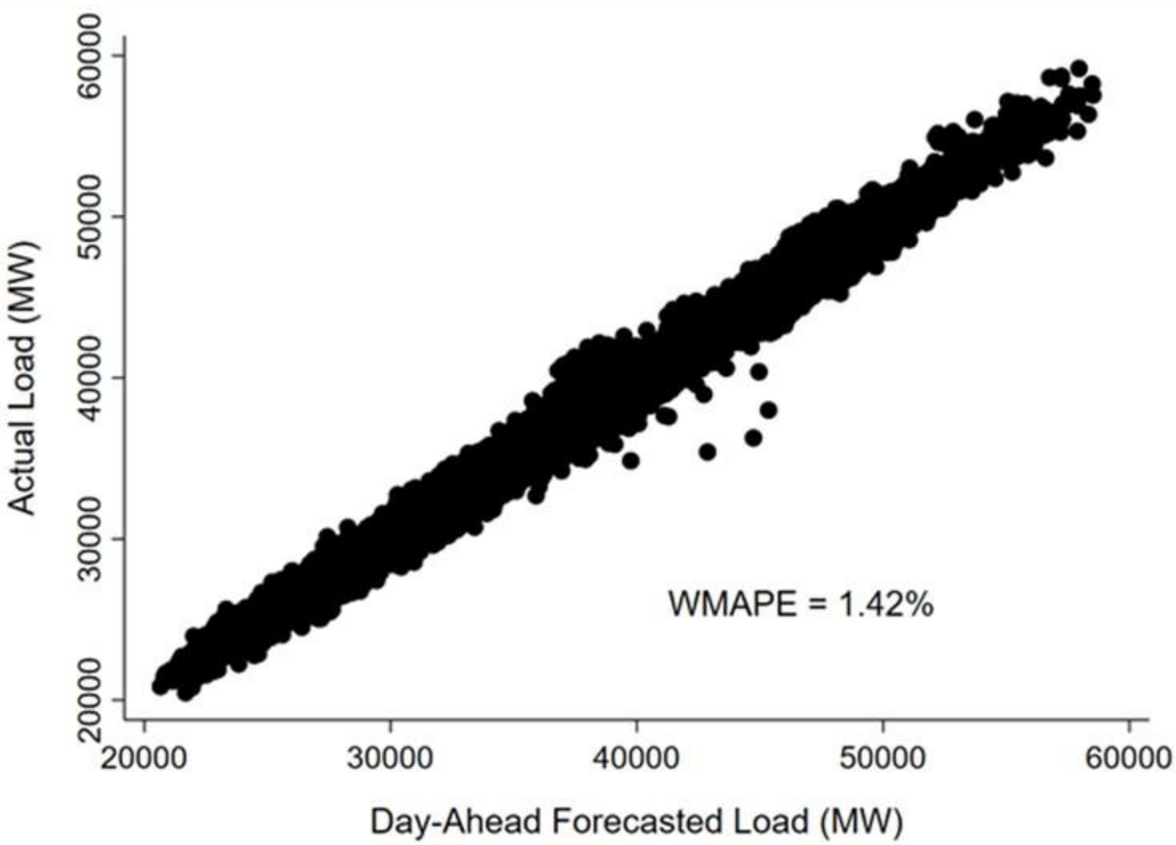

Figure 4: National Grid's day-ahead Load forecast and the realized level of load, 1 Jan 2012–31 Dec 2013

5 How does Wind Energy in Great Britain Compare with Conventional Forms of Generation?

In Great Britain, generating stations inform the system operator of their intended level of generation one hour prior to real-time. This value is known as the final physical notification (FPN). Generators also submit bids (proposals to reduce generation) and offers (proposals to increase generation) to provide balancing services. During real-time, the system operator accepts bids and offers based on its assessment of system conditions. The revised generation schedule equals the FPN plus the level of balancing services volume requested by the system operator.

Failure to follow the revised generation schedule gives rise to an electricity market imbalance that needs to be resolved by other generators.

Using FPN and balancing services data from National Grid, we compared scheduled levels of generation with metered generation. For CCGT, the WMAPE is 4.18% (Figure 5); for nuclear, 5.85% (Figure 6); for coal-fired generating stations, 1.89% (Figure 7); and for wind energy, 13.84% (Figure 8).

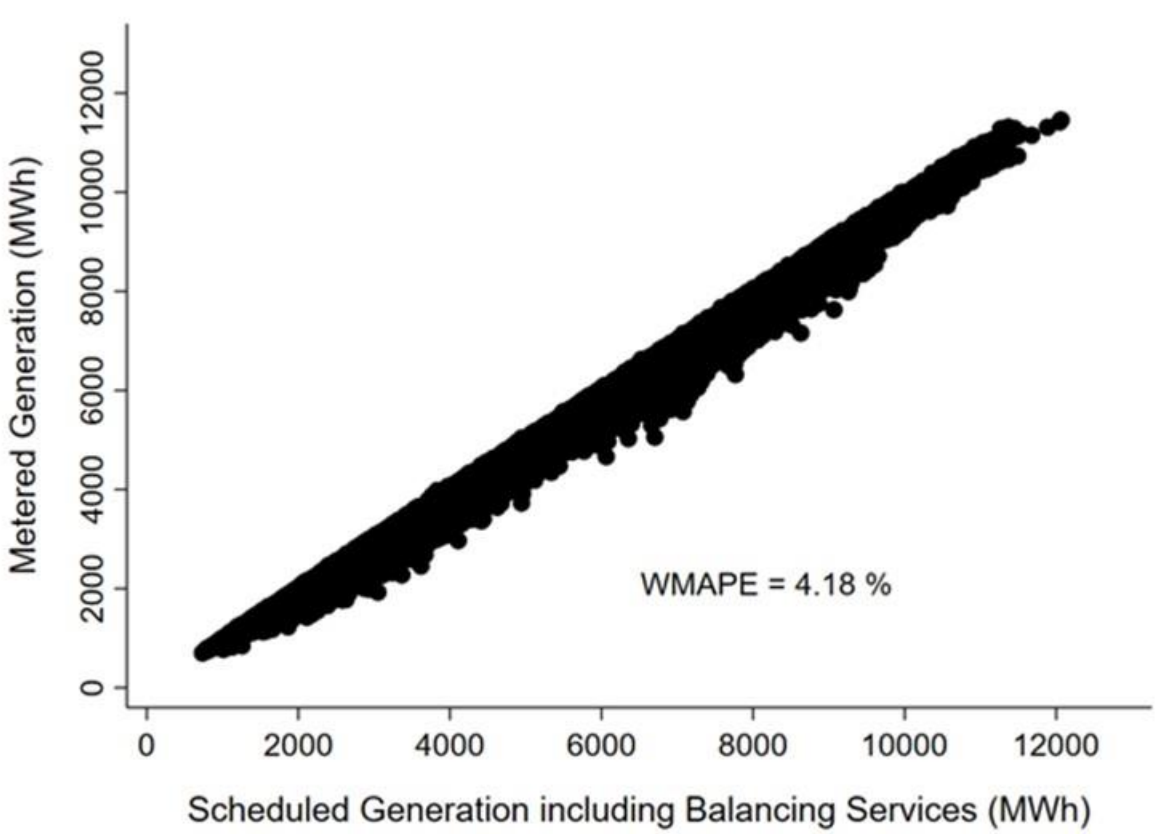

Figure 5: Actual vs. scheduled generation: The case of CCGT in Great Britain, 1 Jan 2012–31 Dec 2013

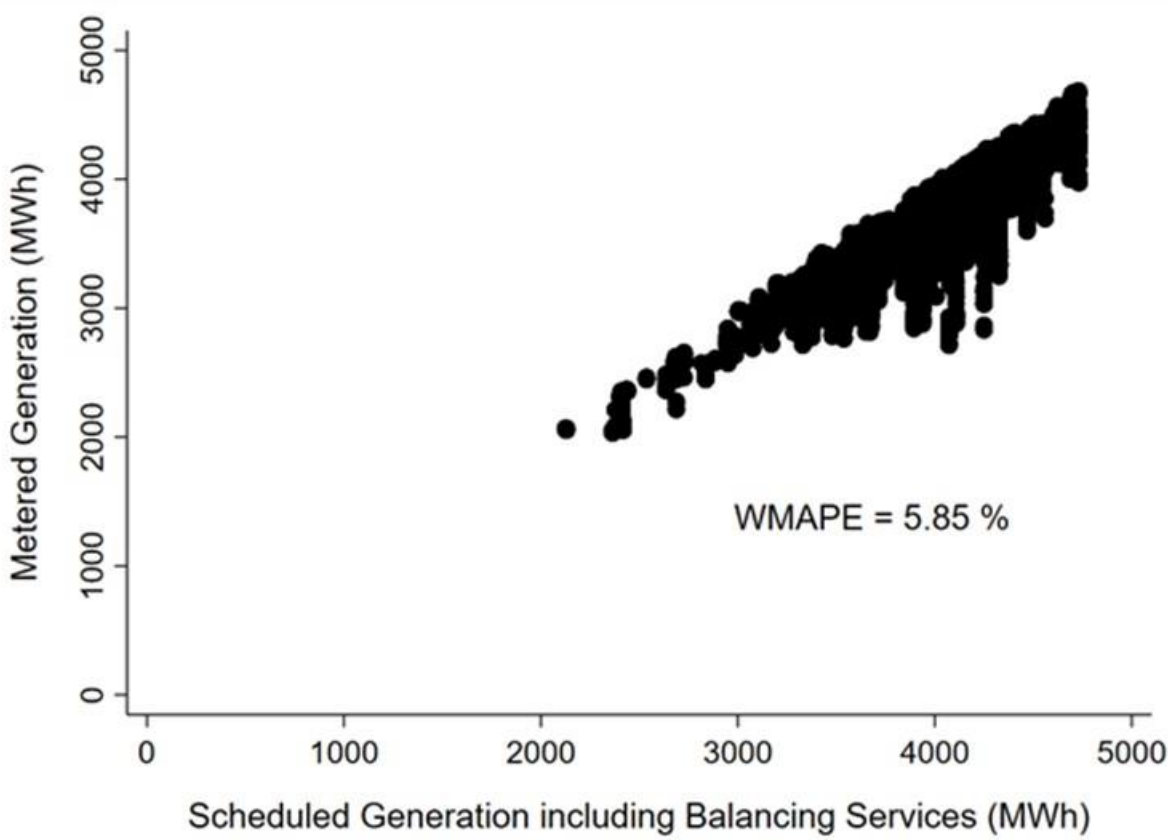

Figure 6: Actual vs. scheduled generation: The case of nuclear energy in Great Britain, 1 Jan 2012–31 Dec 2013

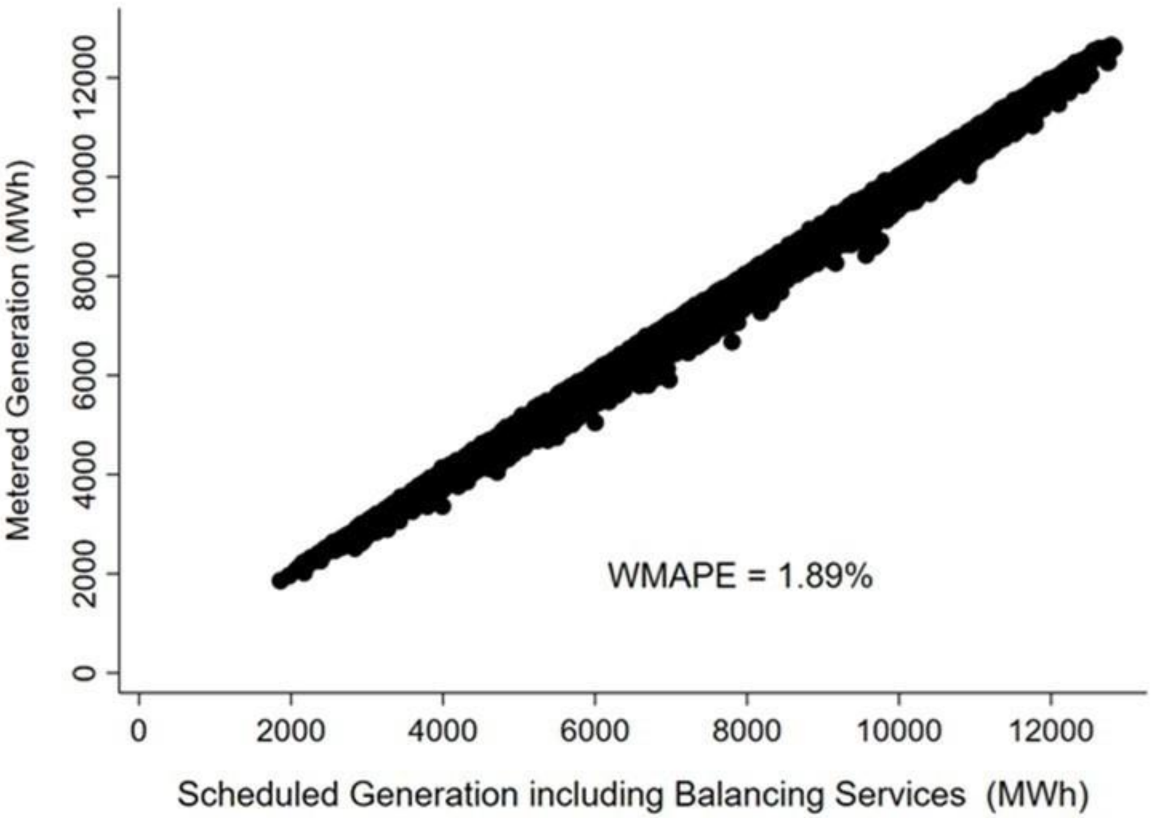

Figure 7: Actual vs. scheduled generation: The case of coal-fired generation in Great Britain, 1 Jan 2012–31 Dec 2013

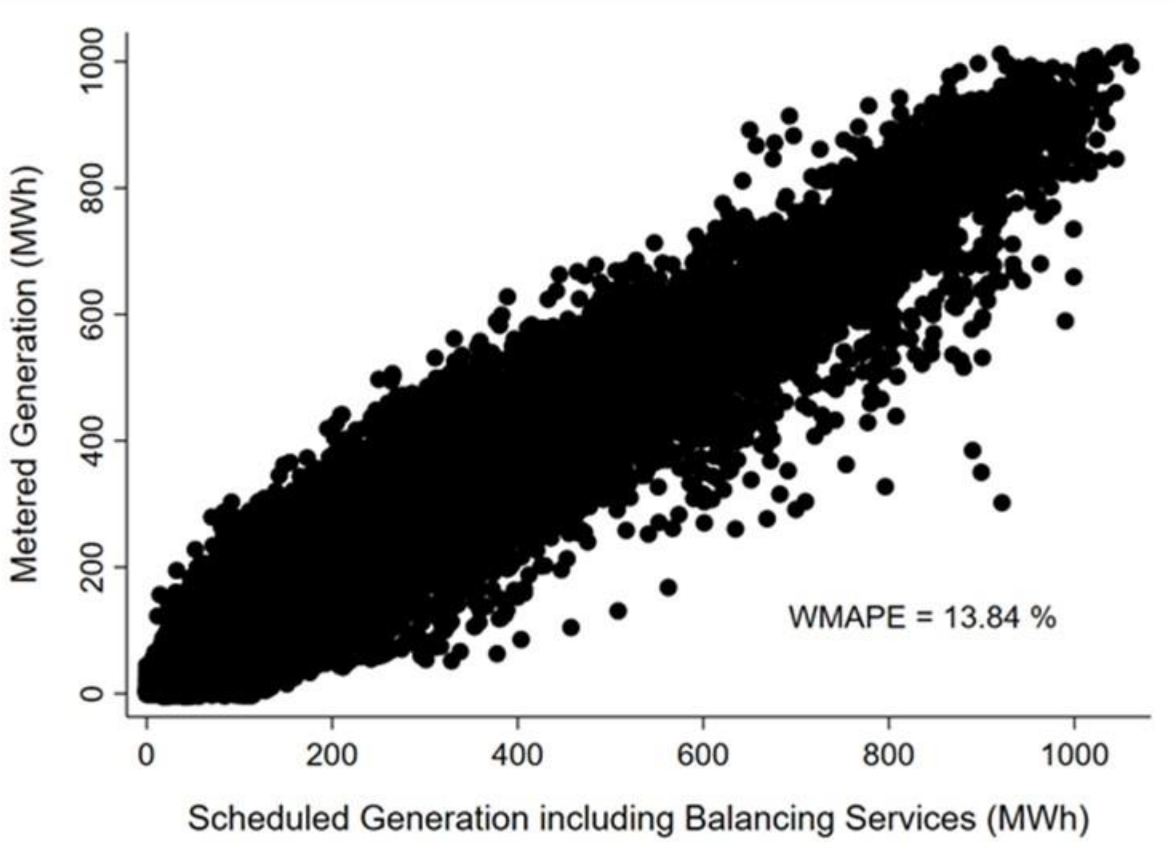

Figure 8: Metered vs. scheduled generation: The case of wind energy generation in Great Britain, 1 Jan 2012–31 Dec 2013

The data from National Grid also make it possible to calculate energy imbalances by fuel type. Based on these data, we report the mean levels of imbalances weighted by the mean levels of metered generation in Figure 9. The energy imbalances for coal, CCGT, and nuclear are substantially less than 10%, while the imbalance that National Grid attributes to wind energy exceeds 30%. This suggests that a substantial portion of the energy scheduled to be produced using wind turbines is actually generated using some other energy source. This is consistent with Wheatley's finding (2013) that the per MWh reductions in CO2 attributable to wind energy in Ireland is less than the implied average carbon intensity in the absence of wind energy.   It is also consistent with Forbes and Zampelli (2019) who report that CO_2 benefits of wind energy penetration given current operational methods are subject to diminishing marginal returns.

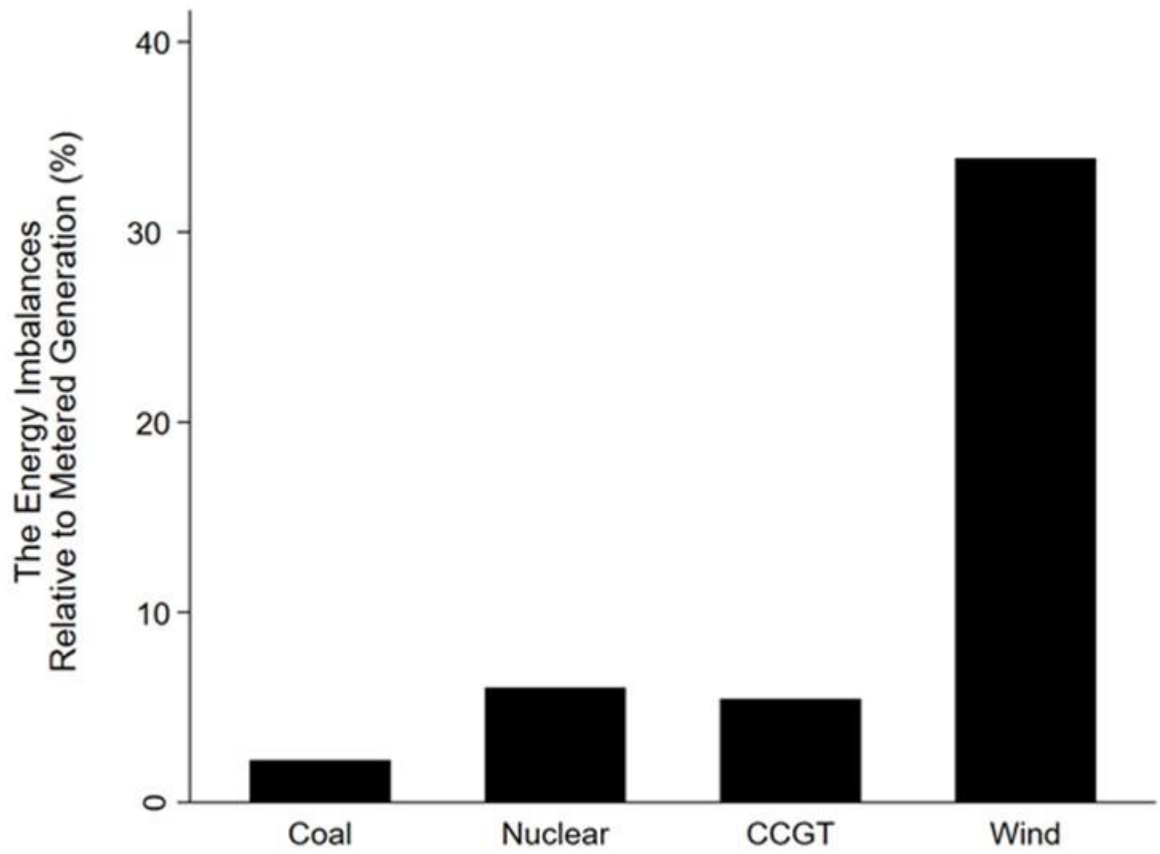

Figure 9: Energy Imbalances by Fuel Type, 1 Jan 2012–31 Dec 2013

A critic has asserted that the wind energy WMAPE for the period 1 January 2012 through 31 December 2013 is not representative. To assess this claim, we obtained the half-hour metered generation, FPN, and balancing services data for wind energy stations directly connected to the transmission system in Great Britain for the period 1 April 2005 through 31 December 2011. The calculated value of WMAPE for this period is 22.64% (Figure 10). In this sense, the results for the 2012-2013 period represent an improvement.

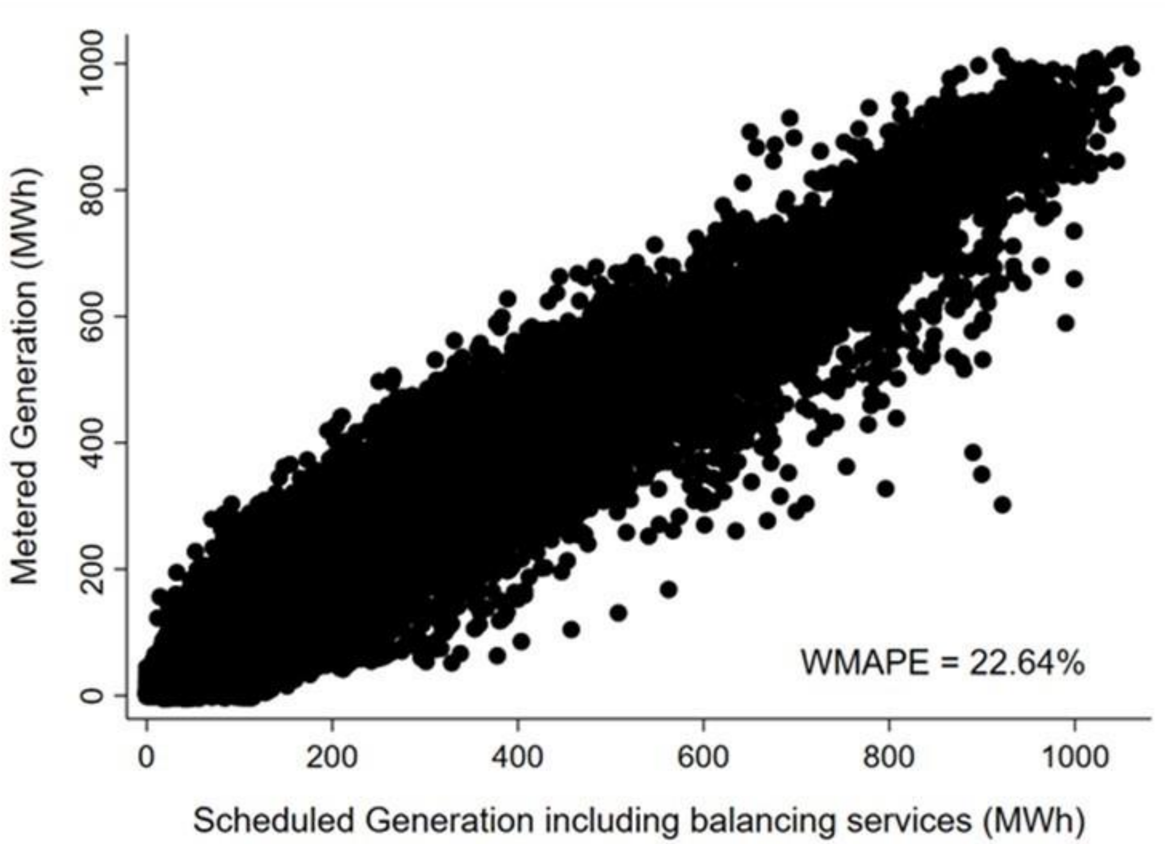

Figure 10: Metered vs. scheduled generation for wind energy generating stations in Great Britain directly connected to the transmission system, 1 Apr 2005–31 Dec 2011

6 Prospects for Improving Forecast Accuracy

Improving forecast accuracy requires as a first step an examination of the errors in the existing forecasts. We begin by pointing out that the errors in an optimal day-ahead forecast should be uncorrelated with information that is known prior to real-time. In the case of a day-ahead forecast, the error should be uncorrelated with the errors in previous days. Moreover, the errors should uncorrelated with the day-ahead level of forecasted wind energy given that this is obviously known when the day-ahead forecast is generated. With respect to meteorological conditions, it follows that while the errors in an optimal forecast will most likely be correlated with unexpected meteorological conditions, they will not be correlated with meteorological conditions that that are expected to occur.

With the notion of the errors of an optimal forecast in mind, we proceed by first reporting that the wind energy forecast errors are characterized by a systematic time-series component (Figure 11). Observe that there is a nontrivial correlation in the errors beyond one day, i.e. beyond 48 lags (equivalent to one day given that there are 48 market periods in each day).

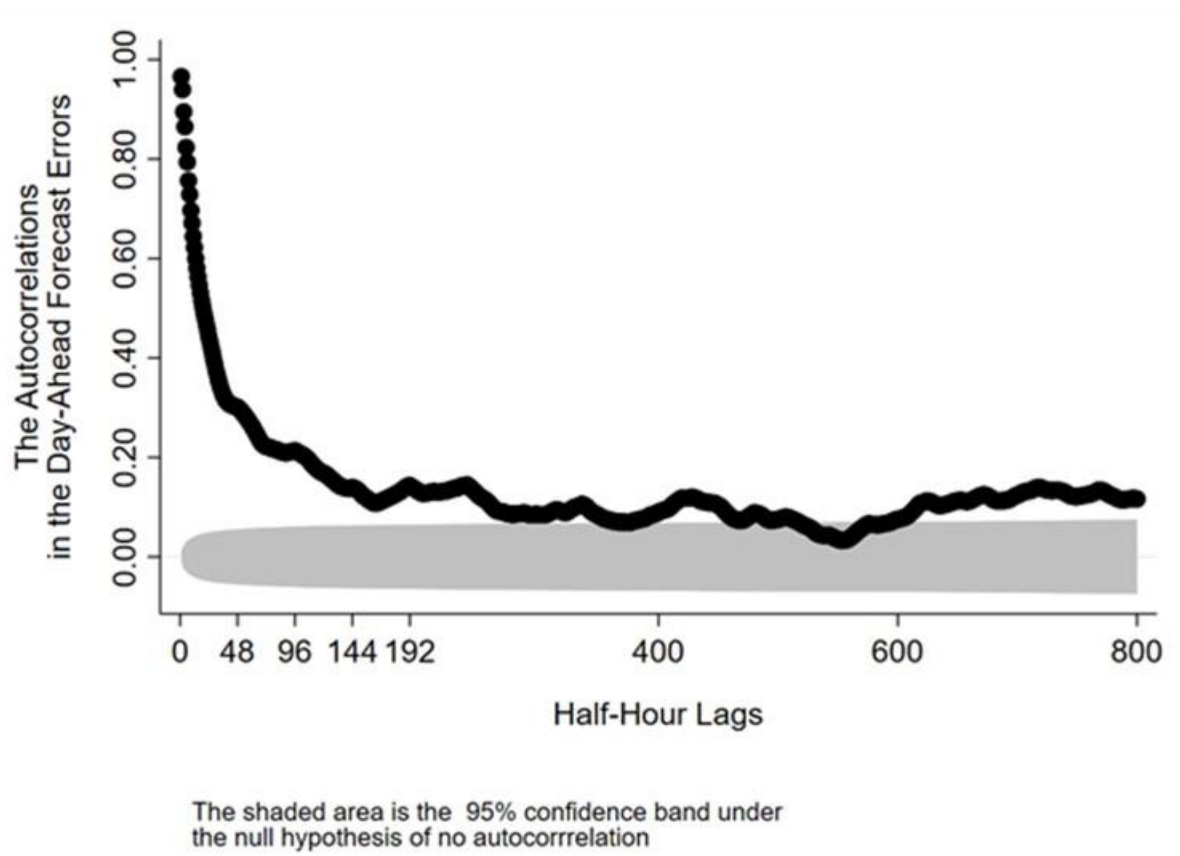

Figure 11: The Autocorrelations in the Day-Ahead Wind Energy Forecast Errors, 1 Jan 2012–31 Dec 2013

We then assess the relationship between the day-ahead wind energy forecast errors and forecasted weather conditions. As Scotland is the dominant source of wind energy in Great Britain, the analysis makes use of an archive of hourly day-ahead weather forecasts for Edinburgh Airport in Scotland. We obtained these data from CustomWeather, a specialized provider of weather forecasts based in San Francisco (http://customweather.com/).

Ordinary least squares (OLS) regression of National Grid's wind energy forecast errors on day-ahead forecasted weather conditions for Edinburgh yields highly statistically significant coefficients for forecasted temperature, dewpoint, visibility, and precipitation probability (Table 1). For example, the OLS results indicate that the error in the wind energy forecast (measured in MW) is higher, the higher the forecasted wind speed measured in kilometers per hour. This is evidence that there is a flaw in the modeled relationship between forecasted wind speed and forecasted wind energy. The OLS analysis of the errors results also indicate that wind energy forecast errors are not independent of forecasted wind energy levels. The finding is not unique to National Grid. For example, Forbes and Zampelli (2017) have reported that OLS analysis of wind forecast errors in the 50Hertz control area in Germany yields similar results. In our opinion, this shortcoming is one of the consequences of choosing a metric of forecast accuracy that mistakenly telegraphs users that the forecasts are accurate.

Table 1: Parameter estimates for wind energy forecast errors in Great Britain

Variable	Estimated Coefficient	T Statistic	P Value
Constant	-712.676	-9.27	< 0.001
ForecastedWindEnergy	-0.15	-68.43	< 0.001
ForecastedTemp	11.942	5.42	< 0.001
ForecastedWindSpeed	9.642	33.35	< 0.001
ForecastedHumidity	7.519	9.56	< 0.001
ForecastedDewPoint	-20.361	-8.55	< 0.001
ForecastedVisibility	6.397	4.78	< 0.001
ForecastedProbPrecip	-1.234	-7.29	< 0.001
R-Square (OLS)	0.169		
Number of observations	29,734		

Note: The analysis excludes those market periods in which wind energy output was affected by the provision of wind energy balancing services to the system operator.

7 A Model to Enhance Forecast Accuracy

Based on the reported OLS findings, we formulate a model to predict wind energy generation as a function of forecasted wind energy generation, the FPNs of wind energy operators, and forecasted weather conditions. The model includes binary variables for hour of day, season, and year. The linear version is given by:

$$\begin{aligned}
\mathit{WindEnergy} &= \mathit{Constant} + \alpha_1\, \mathit{WindFPN} + \alpha_2\, \mathit{DaForecastedWindEnergy} \\
&+ \alpha_3\, \mathit{DaForecastedTemp} + \alpha_4\, \mathit{DaForecastedWindSpeed} + \alpha_5\, \mathit{DaForecastedHumidity} \\
&+ \alpha_6\, \mathit{DaForecastedDewPoint} + \alpha_7\, \mathit{DaForecastedVisibility} + \alpha_8\, \mathit{DaForecastedProbPrecip} \\
&+ \sum_{i=2}^{48} \phi_i \mathit{PeriodofDay}_i + \sum_{j=2}^{61} \gamma_i \mathit{Season}_j + \delta\, \mathrm{Year2012} + \varepsilon \qquad (7)
\end{aligned}$$

where

WindEnergy is the metered level of wind energy generation in MW corrected for the provision of balancing services;

DaForecastedTemp is the day-ahead forecasted temperature measured in Kelvin;

DaForecastedWindSpeed is the day-ahead forecasted wind speed;

DaForecastedHumidity is the day-ahead forecasted relative humidity;

DaForecastedDewPoint is the day-ahead forecasted dewpoint measured in Kelvin;

DaForecastedVisibility is the day-ahead forecasted visibility;

DaForecastedProbPrecip is the forecasted probability of precipitation represented as a percent.;

DaForecastedWindEnergy is the day-ahead forecasted level of wind energy generation measured in MW;

WindFPN is the final physical notification of expected generation by the wind energy generating stations;

PeriodofDay is a series of binary variables representing the 48 electricity market periods of each day;

Season is a series of binary variables representing the season; Exclusive of the leap year in 2012, each of these binary variables represents six consecutive days. Thus, the variable represents 61 "seasons" over the course of a year.

Year2012 is a binary variable that is equal to one if year equals 2012;

ε is the random error.The model was estimated using a sample of 33,822 observations for the period 1 January 2012 through 31 December 2013. To test whether the assumption of linearity is justified, we follow Box and Cox (1964) and transform the dependent variable as

$$TWindEnergy = \begin{cases} (WindEnergy + \lambda_2)^{\lambda_1} - 1)/\lambda_1 \,; \lambda_1 \neq 0 \\ \ln(WindEnergy + \lambda_2) \,;\ \lambda_1 = 0 \end{cases} \tag{8}$$

where:

λ_1 is a parameter estimate obtained from the Box-Cox procedure;

λ_2 is a value that ensures that the left-hand side of equation (7) is positive, in this case, one MW.

Inspection of (7) and (8) reveals that the Box-Cox transformation preserves the direction of the relationships that the original dependent variable has with the explanatory variables. Under the assumption of linearity in the explanatory variables, the estimate of λ_1 is 0.599 with a *p*-value < 0.001. Linearity in the dependent variable is thus strongly rejected. To test the assumption of linearity in the explanatory variables, we relied on the multivariable fractional polynomial (MFP) methodology, a useful technique when one suspects that some of the relationships between dependent and explanatory variables might be nonlinear (Royston and Sauerbrei, 2008). The approach begins by estimating a model strictly linear in the explanatory variables. Subsequent estimations cycle through a battery of nonlinear transformations of explanatory variables (e.g., cube roots, square roots, squares, etc.) until the model that best predicts the dependent variable is found.

In the current context, results provided support for specifying some of explanatory variables with powers other than unity. These estimates, however, are conditional on the estimated value of the Box-Cox parameter λ_1 so that inconsistencies between the two cannot be ruled out. Fortunately, the solution is straightforward—iterate between the two until the Box-Cox transformation of the dependent variable is consistent with the MFP transformations of the explanatory variables. The process yielded a revised estimate of 0.55497 for λ_1 and a transformed structural equation given by:

$$
\begin{aligned}
&\mathit{TWindEnergy} = \mathit{Constant} + \beta_1\, \mathit{WindFPN}^{2/3} + \beta_2\, \mathit{WindFPN} + \beta_3\, \mathit{DaForecastedWindEnergy}^{1/3} \\
&+ \beta_4\, \mathit{DaForecastedTemp} + \beta_5\, \mathit{DaForecastedWindSpeed}^{3} \\
&+ \beta_6\, \mathit{ln(DaForecastedWindSpeed)}{*}\mathit{DaForecastedWindSpeed}^{3} + \beta_7\, \mathit{DaForecastedHumidity}^{3} \\
&+ \beta_8\, \mathit{DaForecastedDewPoint} + \beta_9\, \mathit{DaForecastedVisibility}^{0.5} + \beta_9\, \mathrm{DaForecastedProbPrecip}^{3}
\end{aligned}
$$

$$+ \sum_{i=2}^{48} \phi_i PeriodofDay_i \quad + \quad \sum_{j=2}^{61} \phi_i Season_j \tag{9}$$

Though least squares estimation of (9) produces a seemingly impressive R^2 of about 0.97, the parameter estimates are highly suspect given that a Portmanteau (*Q*) test (Box and Pierce, 1970; Ljung and Box, 1978) indicates that the OLS residuals are not "white noise." The error distribution also exhibits greater kurtosis, i.e., a higher incidence of outcomes in the tails, than the Gaussian distribution. Not surprisingly, the null hypothesis of normality is rejected by the Shapiro-Wilk Normality Test (Shapiro & Wilk, 1965).

We address the problem of autocorrelation in the residuals using an autoregressive–moving-average with exogenous inputs (ARMAX) model comprised of two components. The first is a structural component equivalent to the Box-Cox/MFP transformed equation discussed above, the second an ARMA component which models the dependency of the wind energy generation at time t on wind energy generation in previous time periods as a combination of an AR(*p*) process and a MA(*q*) process. The modeled lag lengths are $p, q = 1$ through 11. Selected parameter estimates are reported in Table 2.

The conditional variance was not assumed to be constant. Instead, the conditional variance was modeled as a function of 61 variables. These variables include transformed measures of the FPNs, forecasted wind speeds, forecasted probability of precipitation, binary measures of seasonality and the hour of the day. Thirty-one of these variables were determined to be statistically significant.

The error term was not assumed to be Gaussian. Instead, the distribution of error term was modeled using the Student t distribution which allows for more kurtosis. Based on Harvey (2013,

p. 20) and the degree of freedom parameter reported by the estimation, the model accommodates a kurtosis level that is about twice the level recognized by the Gaussian distribution.

Table 2: Estimation results

Variable	Estimated Coefficient	Standard Error	T Statistic	P Value
Const	68.3224	16.7827	4.07	< 0.001
$WindFPN^{0.667}$	0.1013	0.0259	3.91	< 0.001
WindFPN	0.0004	0.0014	0.27	0.785
$DaForecastedWindEnergy^{0.3333}$	0.8442	0.0845	9.99	< 0.001
ForecastedTemperature	-0.0540	0.0457	-1.18	0.237
$ForecastedWindSpeed^3$	0.0003	0.0001	3.59	< 0.001
$Ln(ForecastedWindSpeed)*ForecastedWindSpeed^3$	-7.6E-05	2.2E-05	-3.53	< 0.001
$ForecastedHumidity^3$	-3.1E-06	1.1E-06	-2.7	0.007
ForecastedDewPoint	0.1730	0.0758	2.28	0.023
$ForecastedVisibility^{0.5}$	-0.1466	0.1029	-1.42	0.154
$ForecastedProbPrecip^3$	1.0E-06	9.9E-07	1.05	0.296
Year2012	-23.0338	5.1671	-4.46	< 0.001
R-Squared based on all estimated parameters	0.995			
R-Squared based on the estimated structural parameters exclusively	0.472			
Number of Observations	33,822			
28 of the binary variables representing the market period of the day are statistically significant. 33 of the binary variables representing the season of the year are statistically significant. With respect to the time-series variables, 9 of the AR terms and 8 of the MA terms are statistically significant. The specific parameter estimates are not reported here but are available from the corresponding author.				

Observe that the estimated coefficient for *ForecastedDewPoint* is nontrivial in magnitude and statistically significant. Other statistically significant continuous variables include *$DaForecastedWindEnergy^{0.3333}$*, *$WindFPN^{0.667}$*, *$ForecastedWindSpeed^3$* and *$Ln(ForecastedWindSpeed)*ForecastedWindSpeed^3$, and $ForecastedHumidity^3$*. Twenty-eight of binary variables representing unobserved market period specific effects and 33 of the variables that represent unobserved seasonal effects are statistically significant. The R^2 based only on the structural parameters of the model equals 0.472 while the R^2 based on all the parameter estimates including the ARMA terms equals 0.995. The differential in these two values is a clear indication of the importance of the ARMA terms to the model's explanatory power.

The model's predictive power was evaluated for the period 1 January–30 June 2014. The WMAPEs, CWRMSEs, and EWRMSEs for the out-of-sample period are reported in Table 3. To provide perspective on the values reported in the table, the WMAPE on the day-ahead load forecast over this period was 1.46 %, while the EWRMSE was about 1.89 %. Based on the 30-minute peak level of electricity consumption over the period 1 January 2012–30 June 2014, the capacity of the equipment powered by electricity was taken to be 59,199 MW. Based on this value, the CWRMSE for the load forecast was about 1.13%.

Over the out-of-sample evaluation period, the day-ahead wind energy forecasts have a WMAPE of 23.65 %. The FPN-based projections yield a WMAPE of about 21.31% (Figure 12). In contrast, model-based forecasts yield period-ahead wind energy forecasts with an WMAPE of about 3.41% and predictive R^2 of 0.995. Period-ahead (period t) predictions, however, may be of limited value to system operators as they are available only at the end of the previous period (period t-1). Accordingly, we calculated predictions that could be made available to a system operator one full period prior to real-time, i.e., at the end of period t-2. For this case, the out-of-sample WMAPE is 5.76% (Figure 13). A similar pattern is evident if one measures the error using the EWRMSE or the CWRMSE.

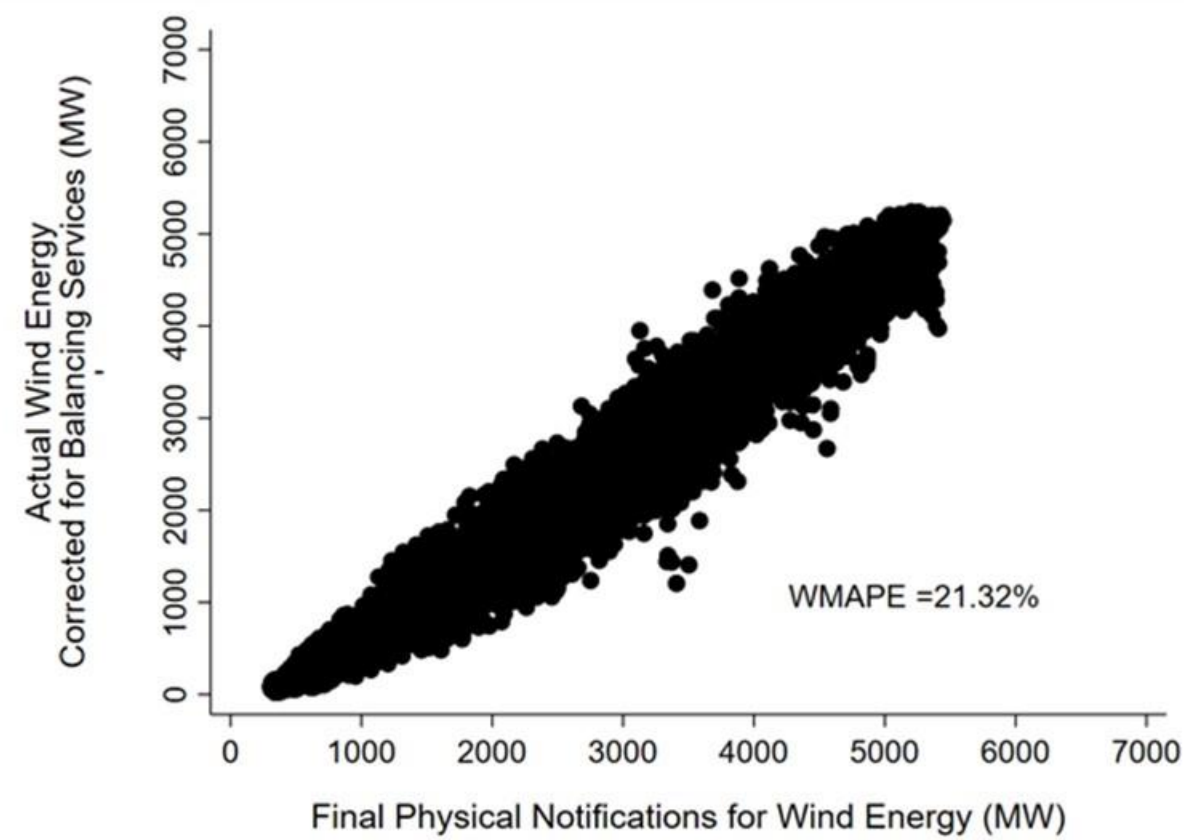

Figure 12: The final physical notifications by wind energy operators to the system operator and the realized level of wind energy generation corrected for balancing actions, 1 Jan–30 June 2014

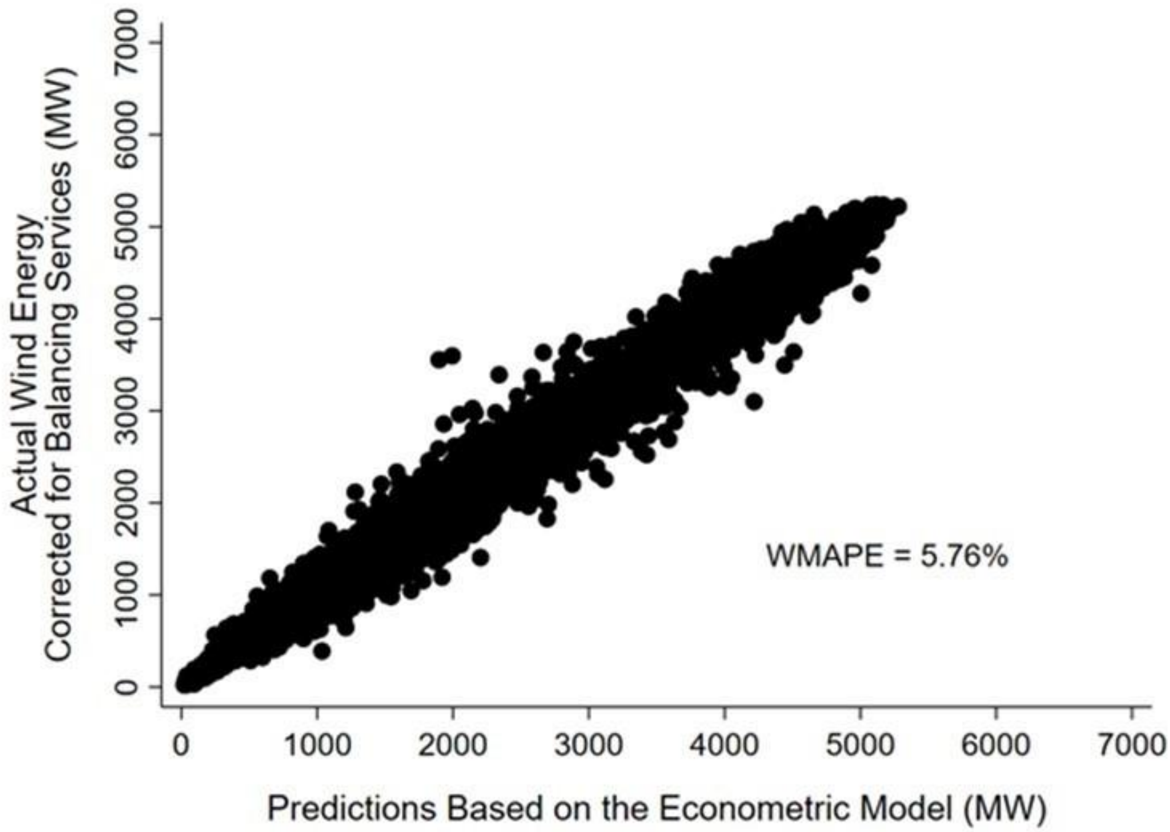

Figure 13: One period-ahead out-of-sample econometric predictions and the metered level of wind energy generation corrected for balancing actions, 1 Jan–30 June 2014

Table 3*:* Out-of-sample results for wind energy forecasting in Great Britain, 1 Jan 2014 – 30 June 2014

	WMAPE %	EWRMSE %	CWRMSE %
Day-Ahead Forecast	23.65	31.98	9.79
Forecast equal to the generation level declared by operators one hour prior to real-time.	21.31	24.32	7.45
Econometrically Modified one-full period-ahead Forecast	5.76	8.12	2.49
Econometrically Modified Period-Ahead Forecast	3.41	4.77	1.46

8 Summary and Conclusion

This paper successfully challenges the claim that wind energy forecasts in Great Britain are highly accurate. The claim, based on normalizing the forecast error by wind energy "capacity," conflates the forecasting error in term of energy with the capacity factor. Because the average capacity factor for wind energy in Great Britain is relatively low, the capacity weighted error is modest, which in turn gives the mistaken impression that wind forecasts are highly accurate. Using a metric consistent with the guidance of professional forecasters who work outside the wind energy industry offers evidence that day-ahead wind energy forecasts in Great Britain are much less accurate than day-ahead load forecasts. We also presented evidence that energy imbalances associated with wind energy in Great Britain are far larger relative to metered generation than the imbalances associated with other energy sources.

Regarding prospects for forecast improvement, we demonstrated that the errors in the wind energy forecasts are not purely random. Specifically, we have shown that day-ahead wind energy forecast errors are not independent of the level of forecasted wind energy or day-ahead forecasted weather conditions. We have shown that the wind energy forecast errors are not "white noise," but are characterized by systematic time-series components. Consequently, we demonstrated how an ARMAX model can yield very short-run wind energy forecasts that are significantly more accurate than existing forecasts. Though this does not fully resolve the challenge posed by wind energy's intermittency, it may enable wind energy to significantly improve its capability to reduce carbon emissions without endangering the reliability of the power grid.

References

Box, G. E. P. and D. R. Cox. (1964). An analysis of transformations. Journal of the Royal Statistical Society, Series B, 26, 211–252.

Box, G. E. P., and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the American Statistical Association 65: 1509–1526.

Cali, U., B. Lange, R. Jursa, and K. Biermann. (2006). Short-term prediction of distributed generation – Recent advances and future challenges, Elftes Kasseler Symposium Energie-Systemtechnik.

Federal Power Commission. (1967) Prevention of Power Failures: A report to the President by the Federal Power Commission. Available at: http://blackout.gmu.edu/archive/pdf/fpc_67_v1.pdf (Accessed 20 August 2018)

Forbes K. F. and E. M. Zampelli. (2019). Wind energy, the price of carbon allowances, and CO2 emissions: Evidence from Ireland, Energy Policy, 133

Forbes K. F. and E. M. Zampelli. (2017). The Accuracy of Wind Energy Forecasts and Prospects for Improvement, *IEEE Xplore*, 14th International Conference on the European Energy Market, https://doi.org/10.1109/EEM.2017.7981986

Forbes K. F., Marco Stampini, and E. M. Zampelli. (2012). Are Policies to Encourage Wind Energy Predicated on a Misleading Statistic? Electricity Journal 25 (3), 42-54

Harvey, A. C. (2013). Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series, Cambridge Univ. Press, New York.

Holttinen, H., P. Saarikivi, S. Repo, J. Ikäheimo, and G. Koreneff. (2006). Prediction Errors and Balancing Costs for Wind Power Production in Finland. Global Wind Power Conference, Adelaide

Hyndman, R. (2006), Another Look at Forecast-Accuracy Metrics for Intermittent Demand, Foresight: The International Journal of Applied Forecasting, 4: 43-46

Intergovernmental Panel on Climate Change. (2012). Renewable Energy Sources and Climate Change Mitigation Special Report of the Intergovernmental Panel on Climate Change.

Joos, M. and Iain Staffell. (2018). Short-term integration costs of variable renewable energy: Wind curtailment and balancing in Britain and Germany, Renewable and Sustainable Energy Reviews 86 (2018) 45–65

Kariniotakis, G. (2006). State of the art in wind power forecasting, 2nd International Conference on Integration of Renewable Energies and Distributed Energy Resources, Napa, California/USA, 4-8 December.

Krauss, C., B. Graeber, and M. Lange, Integration of 18 GW Wind Energy into the Energy Market (2006). Practical Experiences in Germany, in: Workshop on Best Practice in the Use of Short-term Forecasting of Wind Power, Delft.

Lange, Berhard. (2006). Wind power prediction in Germany – Recent advances and future challenges, Zeitschrift für Energiewirtschaft, vol. 30 (1):115-120.

Lange, Bernhard, Kurt Rohrig, Bernhard Ernst, Florian Schlögl, Umit Cali, Rene Jursa, and Javad Moradi. (2006). Wind power prediction in Germany – Recent advances and future challenges. European Wind Energy Conference and Exhibition, Athens (GR).

Lange, B., K. Rohrig, F. Schlögl, U. Cali, U. and R. Jursa. (2007). Wind Power Forecasting, in Renewable Electricity and the Grid, Godfrey Boyle, Ed. Sterling,VA: Earthscan, London, at 95-120.

Lange, Bernhard, Arne Wessel, Jan Dobschinski, and Kurt Rohrig. (2009). Role of Wind Power Forecasts in Grid Integration Kasseler Symposium Energie-Systemtechnik, at 118-130

Kolassa S., and W. Schütz. (2007) Advantages of the MAD/Mean Ratio over the MAPE, Foresight:The International Journal of Applied Forecasting, *Issue 6*, *40-43*.

Ljung, G. M., and G. E. P. Box. (1978). On a measure of lack of fit in time series models. Biometrika 65: 297–303.

Milligan, M., K. Porter, E. DeMeo, P. Denholm, H. Holttinen, B. Kirby, N. Miller, A. Mills, M. O'Malley, M. Schuerger, and L. Soder. (2009). Wind Power Myths Debunked, IEEE Power and Energy, November/December vol 7 no 6, at 89-99.

National Audit Office. (2014). Electricity Balancing Services: Briefing for the House of Commons Energy and Climate Change Select Committee

NERC, (2010) IVGTF Task 2.1 Report: Variable Generation Power Forecasting for Operations, at http://www.nerc.com/files/Varialbe%20Generationn%20Power%20Forecasting%20for%20 Operations.pdf

Renewable Electricity Futures Study (Entire Report) National Renewable Energy Laboratory. (2012). Renewable Electricity Futures Study. Hand, M.M., S. Baldwin, E. DeMeo, J.M. Reilly, T. Mai, D. Arent, G. Porro, M. Meshek, D. Sandor. eds. 4 vols. NREL/TP-6A20-52409. Golden, CO: National Renewable Energy Laboratory.

Royal Academy of Engineering, (2014). Wind Energy: implications of large-scale deployment on the GB electricity system. Available at http://www.raeng.org.uk/publications/reports/wind-energy-implications-of-large-scale-deployment

Royston, P., and W. Sauerbrei. (2008). Wiley series in probability and statistics. In Multivariable model-building: A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. Chichester, UK: John Wiley.

Shapiro, S. S. and M.B. Wilk. (1965). An analysis of variance test for normality (complete samples). Biometrika 52: 591–611.

Sovacool, Benjamin K. (2009). The intermittency of wind, solar, and renewable electricity generators: Technical barrier or rhetorical excuse? Utilities Policy, 17 (3-4) 288-296.

Staffell, Iain and Stefan Pfenninger. (2016). Using bias-corrected reanalysis to simulate current and future wind power output, Energy, 114 1224-1239

United Kingdom Electricity Safety, Quality and Continuity Regulations (2002) Available at: http://www.legislation.gov.uk/uksi/2002/2665/pdfs/uksi_20022665_en.pdf (Assessed on March 22 2019)

Wheatley, J., 2013. Quantifying CO_2 Savings from Wind Power, *Energy Policy*, 63 89-96.