

LETTER • **OPEN ACCESS**

Optimising the use of ensemble information in numerical weather forecasts of wind power generation

To cite this article: J Stanger *et al* 2019 *Environ. Res. Lett.* **14** 124086

View the [article online](#) for updates and enhancements.



LETTER

Optimising the use of ensemble information in numerical weather forecasts of wind power generation

OPEN ACCESS

RECEIVED

3 September 2019

REVISED

1 December 2019

ACCEPTED FOR PUBLICATION

3 December 2019

PUBLISHED

19 December 2019

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

J Stanger¹, I Finney², A Weisheimer^{1,3,4} and T Palmer^{1,3}¹ Atmospheric, Oceanic and Planetary Physics (University of Oxford), Sherringdon Road, OX1 3PU, United Kingdom² Lake Street Consulting Ltd, Oxfordshire, United Kingdom³ National Centre of Atmospheric Science (NCAS), United Kingdom⁴ European Centre for Medium-Range Weather Forecasting (ECMWF), Reading, United KingdomE-mail: isla@lakesideconsulting.co.uk**Keywords:** ensemble forecasting, wind energy, energy trading, renewable energy, numerical weather forecastingSupplementary material for this article is available [online](#)**Abstract**

Electricity generation output forecasts for wind farms across Europe use numerical weather prediction (NWP) models. These forecasts influence decisions in the energy market, some of which help determine daily energy prices or the usage of thermal power generation plants. The predictive skill of power generation forecasts has an impact on the profitability of energy trading strategies and the ability to decrease carbon emissions. Probabilistic ensemble forecasts contain valuable information about the uncertainties in a forecast. The energy market typically takes basic approaches to using ensemble data to obtain more skilful forecasts. There is, however, evidence that more sophisticated approaches could yield significant further improvements in forecast skill and utility. In this letter, the application of ensemble forecasting methods to the aggregated electricity generation output for wind farms across Germany is investigated using historical ensemble forecasts from the European Centre for Medium-Range Weather Forecasting (ECMWF). Multiple methods for producing a single forecast from the ensemble are tried and tested against traditional deterministic methods. All the methods exhibit positive skill, relative to a climatological forecast, out to a lead time of at least seven days. A wind energy trading strategy involving ensemble data is implemented and produces significantly more profit than trading strategies based on single forecasts. It is thus found that ensemble spread is a good predictor for wind electricity generation output forecast uncertainty and is extremely valuable at informing wind energy trading strategy.

1. Introduction

The wind is an important source of renewable energy, and its use is growing. In Germany, wind power generation capacity increased nearly ten-fold between the years 2000 and 2018—from 6 GW to 59 GW—and in 2018 produced 19 % of all the electricity consumed in Germany [1]. The use of wind power will likely continue to increase in many countries in order to help them meet renewable energy targets.

The rising use of wind power poses significant practical and financial challenges in energy markets and national grid systems related to the predictability of the wind [2]. Electricity's unique nature as a commodity that cannot be stored in large quantities means

that supply and demand must be matched in real time. The variability of wind power therefore presents an operational challenge to power system operators and can incur significant management costs [3].

These challenges can be tackled by improving wind power forecasts. Foley *et al* [4] give a review of wind power forecasting methods, and divide them into two categories: statistical methods which forecast the power generation directly based on past observations [5, 6], and physics-based methods which use numerical weather prediction (NWP) models to forecast the wind speed, and then convert that into a power forecast [7–9]. Comparisons between the two have suggested that NWP models offer significant benefits over statistical methods in the medium range (1–7 d)

[10]. In light of this, we use the ensemble NWP forecasts from the European Centre for Medium Range Weather Forecasting (ECMWF) and focus on the medium range.

In recent years, ensemble forecasting has come to the fore, particularly in the medium range where the chaotic nature of the atmosphere places significant limitations on the value of traditional, deterministic, forecasts [11]. In particular, in such a highly nonlinear system, uncertainty in initial conditions will grow quickly so that increasing model resolution has a limited benefit for the maximum prediction horizon. It also means that predicting the uncertainty in a forecast is very challenging. However, for a forecast to be useful, uncertainty information is essential [12]. Ensemble forecasting tackles both of these issues by running a set, or ensemble, of lower resolution forecasts with perturbed initial conditions such that their distribution represents the uncertainty in the initial conditions. The idea is that the distribution of the ensemble of forecasts will also represent the uncertainty in the forecast [13].

The ensemble mean is limited as a point forecast since it is a statistical measure of a set of individual realisations of a physical model; it may not represent a physically likely state of the atmosphere. Attempts have been made to produce better point forecasts from the ensemble which generally involve calibrating the ensemble members over a sample period and taking a weighted mean [10, 14, 15]. In light of the success of these attempts, we consider a different calibration and weighted mean approach which is less computationally expensive.

In this letter we also consider the operational value of ensemble forecasts to a trader on the financial German electricity market (rather than the physical market). A previous study found that the profitability of trading wind energy can be increased by employing a strategy based on skill forecasts [16]. The authors use complex fuzzy logic methods based on probabilistic forecasts to inform trading strategy. We demonstrate that this is not necessary with the probabilistic information in ensemble forecasts; we implement a straightforward trading strategy which exploits the spread-skill relationship of ensemble forecasts.

This letter begins, in section 2, with the details of the data used and the post-processing of it. In section 3 the forecasting methods listed in section 2.5 are compared in terms of skill and operational value. The final section summarises our conclusions.

2. Methods

2.1. The data

We used historic ensemble forecasts—as opposed to hindcast data: the current model cycle run with historic initial conditions—to represent what was available to the power market at the time, covering the period 08/03/2016 to 30/08/2018 (We provide

motivation for this choice in the supplementary material, which is available at stacks.iop.org/ERL/14/124086/mmedia), of wind speed at 100 m elevation [13, 17]. The ensemble contains 51 members (including the unperturbed control run). The forecasts are initialised at midnight each day and are made for every 3 h to a maximum lead time of 6 days, and then every 6 hours to a maximum lead time of 7 days. The ensemble NWP models have a native resolution of 18 km, and the analysis was done on a $1^\circ \times 1^\circ$ longitude/latitude grid. Also available from ECMWF was a high resolution (9 km native resolution), deterministic, forecast of the same quantities which was initialised with the best-guess initial conditions.

Wind speed observations covering the same time and space ranges came in the form of ERA5 reanalysis data. Power observations were compiled from four companies which cover all grid-integrated wind power generation across Germany (see section 5) [18–21].

Finally, for use in converting wind speeds into wind power, we had monthly data for the total wind generation capacity across Germany, and (assumed constant) fractional capacities contained in each grid cell. All data was divided into *in-sample* and *out-of-sample* data. All analyses were done on the in-sample data and the out-of-sample data was used for testing in order to avoid over-fitting to the data. The in-sample data comprised the even-numbered days compared to a datum and vice-versa (The ‘day’ of a forecast is the day it is made for, not the day it is initialised).

2.2. Conversion to a power forecast

A power forecast is given as a *load factor*—generated power as a fraction of generation capacity. Conversion of the wind speed forecast into a power forecast combines the manufacturer power curve method, used by e.g. Taylor *et al* [10], with the stochastic method, employed by e.g. Sanchez [22]. This project differs from other studies at this point because we calculate aggregated wind power across Germany, and without the exact locations of wind farms. It is hoped that this top-down approach, employing a single power curve to model all wind turbines, is a more practical and efficient approach in an operational scenario as it avoids the spatial interpolation of wind speeds employed by e.g. Cannon *et al* [23].

The wind speed, v_i , at each grid point is first fed through a manufacturer’s power curve, $f(v)$, shown in figure 1, which all wind turbines are assumed to follow [24]. This gives a load factor for each grid point, representing the power generated in each grid cell. A weighted sum of these is calculated, where the weights w_i are the fractional capacities in each grid cell, to give a total *model load factor*, ℓ , for all of Germany:

$$\ell = \sum_i w_i f(v_i). \quad (1)$$

In an ideal scenario, the model load factor would equal the observed load factor, but other factors affect

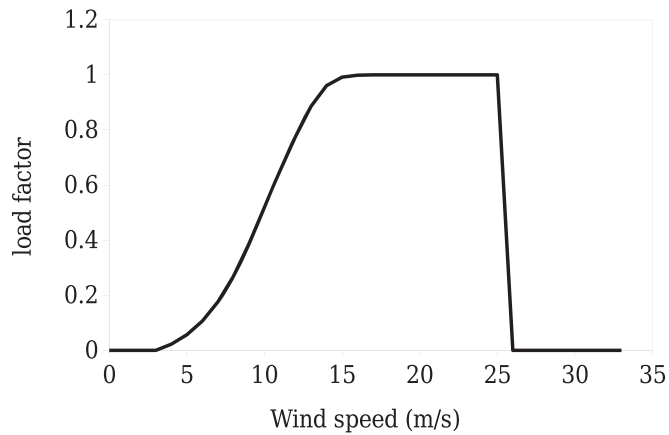


Figure 1. Vesta V90 3 MW manufacturer power curve [24]. The turbine turns for winds exceeding 4 m s^{-1} and cuts out in wind speeds above 25 m s^{-1} . All turbines are assumed to follow this curve.

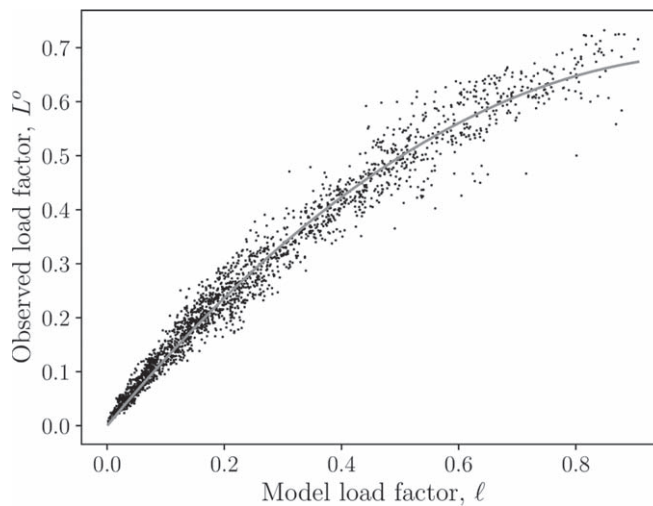


Figure 2. Example power forecast calibration plot with quadratic fit in red, for the winter months (DJF). Solid curve is $y = 1.31x - 0.62x^2$. The tail-off of the curve at high load factors is due to the curtailment of wind power generation.

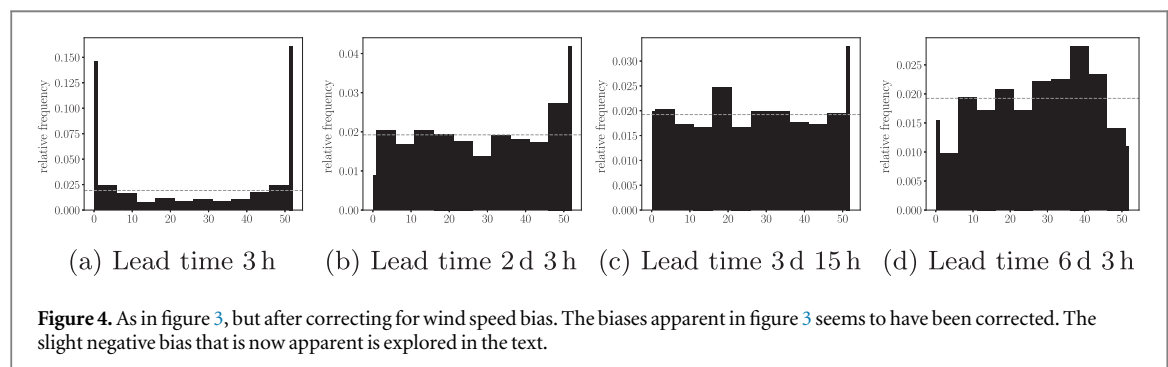
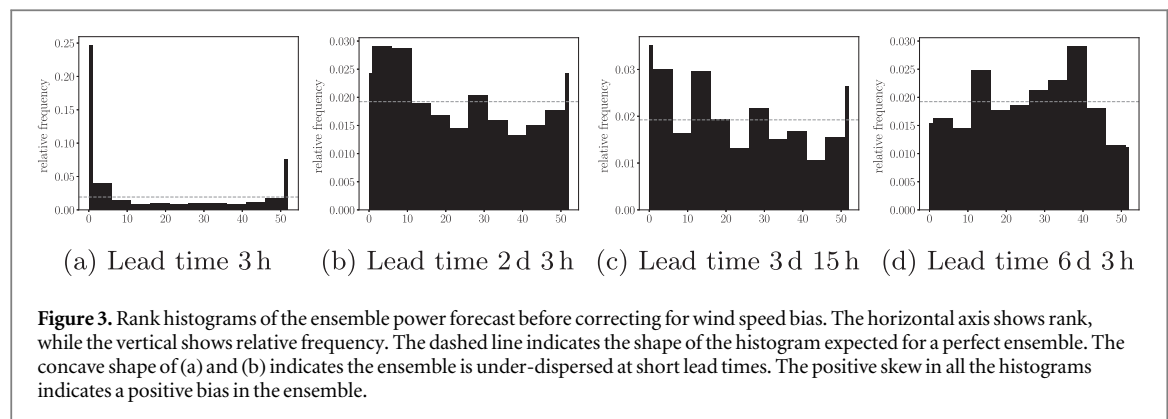
the actual power produced. This can be seen in figure 2, where the model load factor, ℓ , computed for the observed (reanalysis) wind speeds, is plotted against observed load factor, L^o . While the plot is linear for small load factors, it curves downwards at high powers. This is the effect of transmission constraints—if the generated power is too great for the electricity transmission lines to transport to where the demand is then power generation must be curtailed, so power generation is lower than expected. The spread in this plot is largely caused by variation in wind direction and variation in air density. The former is a factor because wind farms are laid out to operate most efficiently in the prevailing wind direction. These two effects were dealt with empirically in our analysis (see below), though the diurnal variation in air density turned out to be observable and was taken into account—see supplementary material for details.

The plot in figure 2 is used to calibrate the model load factor with respect to the true load factor. A curve

of the form $y = ax + bx^2$, was fitted to find parameters a and b . The quadratic term simulates the curtailment discussed above. By computing ℓ for the forecast winds, and passing this through the calibration curve, a calibrated power forecast, $L = a\ell + b\ell^2$, can be computed. The seasonal variation in wind direction and air density was taken into account by using a separate calibration curve for each season (DJF, MAM, JJA and SON).

2.3. Bias correction

The forecast winds must be de-biased with respect to the reanalysis winds in order for the calibration curves to work as described. Rank histograms were made with the ensemble power forecasts and power observations. Rank histograms show the relative frequency of the rank of the observation among the ensemble forecast members [25]. They offer a direct visualisation of the bias and dispersion of the ensemble. The histogram is flat for the ideal ensemble; that is an ensemble where



its members have the same distribution as the quantity being forecast. If the observation consistently falls in the highest or lowest ranks of the distribution, the ensemble has a negative or positive bias, respectively. Similarly, a histogram which is U-shaped implies an under-dispersed ensemble, while an inverted U shape implies an over-dispersed ensemble. For more details, see e.g. Wilks [26]. The rank histograms shown in figure 3 indicate a significant positive bias, particularly at shorter lead times, so forecast biases were corrected.

The mean absolute error (MAE) of the in-sample ensemble mean forecasts was calculated for each lead time with the in-sample data, and this was subtracted off the individual ensemble members. In so doing we assumed that a given ensemble member is unbiased with respect to the ensemble mean. Finally, the bias-corrected forecasts were passed through a filter that set any negative wind speeds to zero.

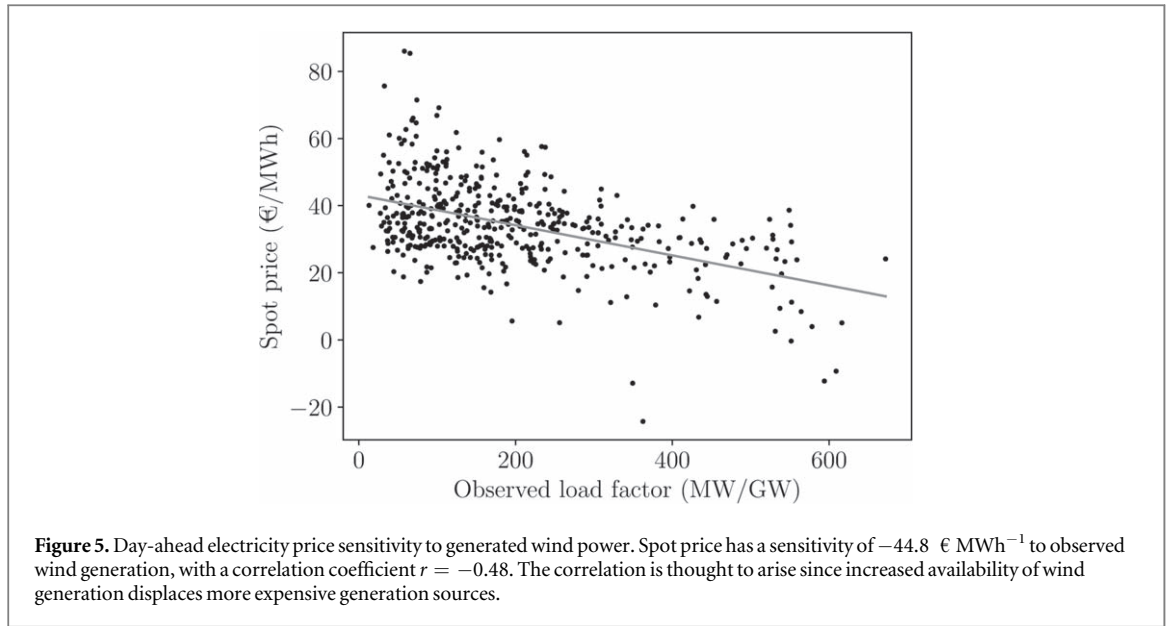
The rank histograms in figure 4 were calculated with bias-corrected forecasts. Much of the positive bias has been removed, though there is a tendency for all the ensemble members to under-predict the power generation in a relatively large number of cases. This is evident from the spikes on the right-hand sides of the histograms in figures 4(b), (c) and (d).

2.4. Trading strategy

Figure 5 shows that the day-ahead price of electricity is sensitive to the forecast amount of wind power generated; the more wind power is generated, the cheaper the electricity. This is because wind power is

less expensive on the market than many other generation sources, partly because of government feed-in tariffs. Wind power availability reduces demand for more expensive sources such as coal or oil fired power stations. This sensitivity can be used by market participants to make a profit. Consider a third party who is bound by a contract to deliver a fixed quantity of electricity tomorrow. Suppose they have a better forecast than the market and it tells them that there will be more wind power tomorrow than the market believes, i.e. that electricity for tomorrow will be cheaper to buy on the same-day market tomorrow than it is to buy on the day-ahead market today. The third party should sell electricity at the higher price today and then buy it back tomorrow at the lower price. If the better forecast says electricity will be more expensive tomorrow, they should do the opposite; buy today and sell it back tomorrow. In either case, the third party delivers the required amount of electricity and makes a profit.

Such a trading strategy was used to test the forecast techniques listed below. We assumed that the correlation of price with observed load factor is perfectly linear, with gradient -44.8 € MWh^{-1} (see figure 5). We then choose a fixed quantity of electricity, or *position*, to trade each day. In the real world, this must be small enough that the trade does not significantly affect the market; 100 MWh is a reasonable number; for us it is just a scale factor. The daily profit made on the T h-ahead market is then:



$$\begin{aligned} \text{€ (profit)} &= (|L^m(t, T) - L^o(t + T)| \\ &\quad - |L(t, T) - L^o(t + T)|) \\ &\quad \times 44.8 \text{ € MWh}^{-1} \times 10 \text{ MWh}, \quad (2) \end{aligned}$$

where $L^m(t, T)$ is the forecast used by the market and $L(t, T)$ is the third party's forecast. Equation (2) has the desired property that if the third party's forecast is better than the market's then they make a profit, and vice-versa.

2.5. Forecast techniques

The first five methods here are included as commonly used techniques to be improved on by ensemble methods. In particular, *climatology* is the baseline we have chosen to exhibit no skill and thus will show when a forecast is worse than useless. Two methods: *best member* and *perfect forecast* are imaginary forecasts that are deduced *a posteriori*, and are included in the profit and loss analysis of section 3.2 to demonstrate the room for improvement on current forecasting methods. A novel method was devised, namely the *weighted mean*. The *weighted position* is a relatively new method to improve trading strategy and is also compared to other forecast methods in section 3.2 [27]. The term *fixed position* is used to refer to the other forecasting methods, excluding the imaginary ones. The forecast techniques employed were as follows:

Climatology. The historical mean observed power for each hour in the year.

High resolution. The deterministic high resolution wind speed forecast is corrected for bias and used to produce a power forecast.

Control run. Same as *high resolution* but using the unperturbed ensemble member.

Ensemble mean (power). Each member of the bias-corrected ensemble is used to produce a power

forecast using the methodology described in section 2.2. The ensemble mean is then taken.

Ensemble mean (wind). The bias-corrected ensemble mean wind speed is used to produce a power forecast.

Best member. The ensemble member that gives the best (i.e. closest to observation) power forecast is chosen *a posteriori* as the forecast.

Perfect forecast. A forecast is constructed *a posteriori* assuming perfect knowledge of future observations. i.e. $L(t, T) = L^o(t + T)$.

Weighted mean. This new method is like *ensemble mean (power)*, but before taking a mean, the members are ranked from lowest to highest and a weight is given to each ensemble member. The weight for the member with rank n is assigned as follows: take the rank histogram for the relevant lead time and find the bin covering the range of ranks that includes n . The weight is the height of this bin. Occasionally the observation lies outside the ensemble. For these cases, the mean difference between the extreme ensemble member and the observation is calculated. i.e. When the observation lies above the ensemble, the relevant difference d_a is between the highest-ranked ensemble member L_{\max} and the observation: $d_a = L^o - L_{\max}$. When the observation lies below the ensemble $d_b = L_{\min} - L^o$. Now the quantities $L_{\max} + \bar{d}_a$ and $L_{\min} - \bar{d}_b$ (where the bar indicates a mean taken over the in-sample data) are assigned weights given by the heights of the far right and far left bins, respectively. The weighted mean of the ensemble members and these two quantities gives the power forecast.

Weighted position. The quantity of energy traded is multiplied by a factor $\text{RMSE}/\sqrt{2}s$ where the

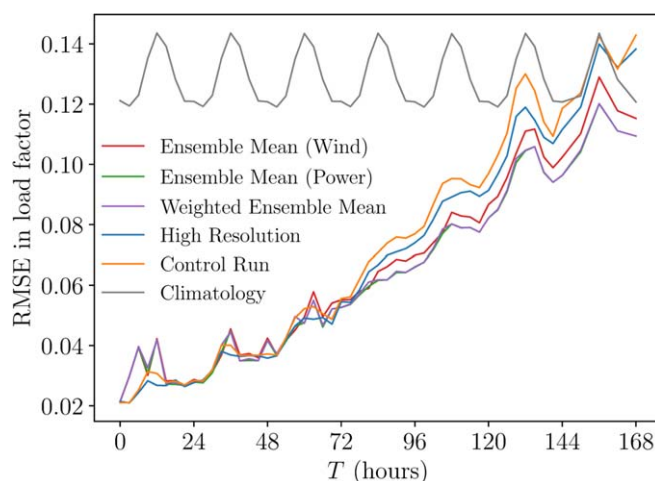


Figure 6. Comparison of the skill of the forecast methods, in terms of RMSE. Methods are largely indistinguishable at short lead times. At longer lead times, ensemble mean methods out-perform the deterministic methods.

RMSE is the root-mean-squared error of an ensemble member and s is the ensemble spread. This is motivated by the spread-skill relationship of ensemble forecasts; it results in a larger (higher-risk) trade when the ensemble spread is smaller. The factor of $\sqrt{2}$ (see derivation in [appendix](#)) normalises the ratio to unity when averaged over a long period to allow for direct comparison with the other trading strategies.

3. Results and discussion

3.1. Forecast skill

For each of the (real) forecast methods listed above the RMSE of the power forecast was calculated for each lead time. Figure 6 shows the results. The RMSEs of the *perfect forecast* and *best member* methods are not shown because they do not meaningfully depend on lead time. For *weighted position*, the RMSE is the same as that of *ensemble mean (power)*. Pleasingly, all NWP forecasting methods are at least as skilful as the climatology out to a lead time of seven days. This is an improvement on the findings by Taylor *et al* [10]. This points to improvements in NWP forecasts over the last 10 years, but also demonstrates that precise wind farm locations are not necessary for skilful forecasts of total countrywide power generation.

At very short lead times (<12 h), the high resolution forecast performs best and the control member also outperforms the various ensemble forecasts. This is thought to be related to the under-dispersion of the ensemble which is evident in figure 4(a); the observation lies outside the ensemble so frequently that we cannot expect the ensemble mean to be particularly skilful.

At longer lead times (more than 3 d) we see a consistent ranking of methods, though it should be noted that the *ensemble mean (power)* and *weighted mean* methods are indistinguishable at all lead times, and so

only *weighted mean* is visible. All the ensemble methods outperform the climatology and deterministic forecasting methods. This reproduces previous findings [10, 28, 29]. The fact that the high resolution forecast out-performs the control is also expected, because the control is run at a lower resolution. *Ensemble mean (power)* out-performs *ensemble mean (wind)* due to the nonlinearity of the power transformation—in particular the manufacturer's power curve. The distinction is particularly important at long lead times when the effect of the nonlinearity is most significant due to the larger ensemble spread (compared to that at shorter lead times). Notable is a diurnal cycle; it is most evident in the climatological forecast, but appears in the others at later lead times. This is explained by the finding (not shown) that the variability in wind power observations is greater during the day than the night.

It is disappointing to see that, in terms of RMSE, *weighted mean* is not measurably more skilful than *ensemble mean (power)*. This is probably because the weightings used rarely differ much from being all-equal. When they do, it is generally the weights for the cases where the observation lies outside the ensemble. This is only one value in a set of 53 and so its effect on the mean is small.

3.2. Trading strategy

A three month period with large fluctuations in power generation was chosen (see figure 7). Figures 8(a) and 9 show profit and loss curves, where the cumulative profit of the trading strategy, described in section 2.4, is plotted over this time period for the forecast techniques listed in section 2.5. The forecast used by the market (L^m) is the climatology, i.e. the climatological forecast would, by construction, make no profit in these plots. Figure 8 also includes the daily profits and losses for the same period for selected methods. The profit and loss curves shown are for the same lead

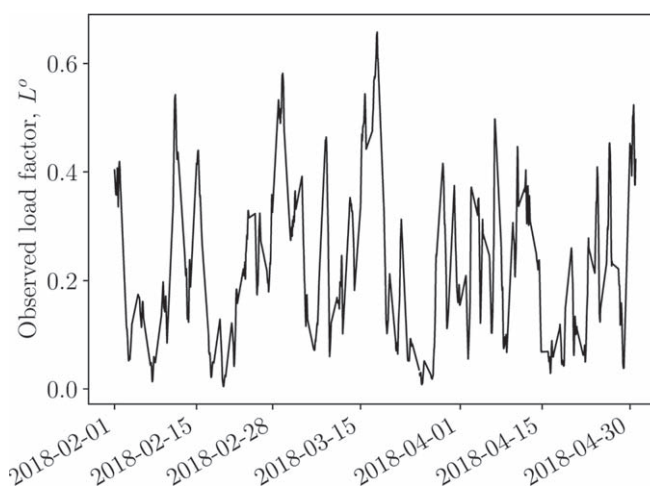
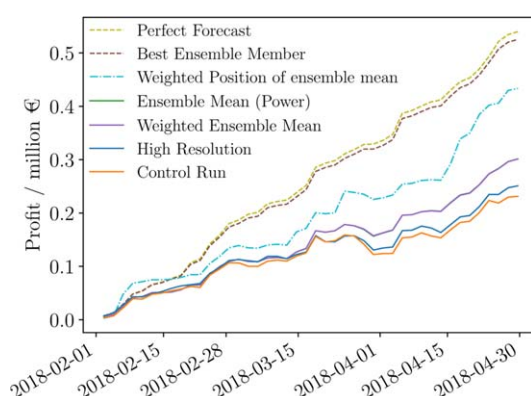
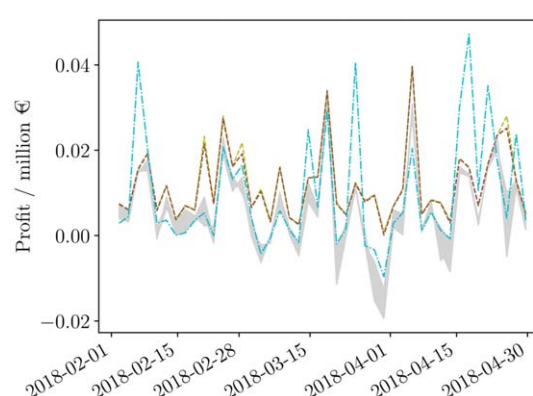


Figure 7. Observed load factor over the period used in the profit and loss analysis. This period was chosen for its large variability in wind power generation.

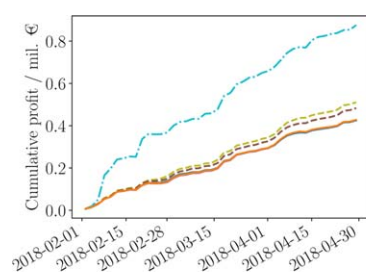


(a) Cumulative profit and loss

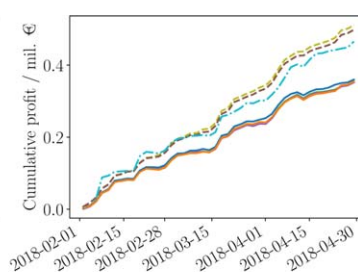


(b) Daily profit and loss for *perfect forecast*, *best member* and *weighted position* (see Figure 8a for legend). The grey shaded region represents the range produced by the fixed-trade methods.

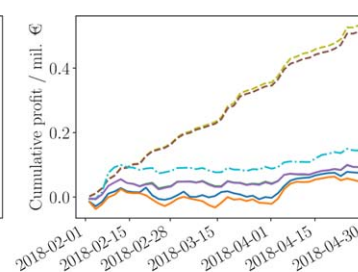
Figure 8. Cumulative and daily profit and loss curves for lead time 3 d 15 h. Daily profit and loss is just the derivative of the cumulative plot. The *weighted position* method successfully identifies occasions when the atmosphere is predictable (when the shaded region in (b) is narrow) and so a proportionally larger trade is made. This means it out-performs all fixed-position-based trading methods.



(a) Lead time 3 h



(b) Lead time 2 d 3 h



(c) Lead time 6 d 3 h

Figure 9. Profit and loss curves at various lead times. Here, the market uses the climatological mean forecast. The ranking of the methods is largely consistent with the ranking of the forecast methods in terms of forecast skill; see figure 6. The *weighted position* method consistently out-performs all the fixed-position methods. Its erroneously large profit in (a) is because the mean position is significantly greater than 100 MWh due to the under-dispersion of the ensemble. See figure 8(a) for the legend.

times as the rank histograms in figures 3 and 4. Note that, at all lead times, profits for all methods are positive, confirming again that all numerical forecasting methods are more skilful than the climatology out to at least seven days.

The relative performance of the forecasting methods largely respects that seen in the RMSE plot in figure 6. At a lead time of 3 h (figure 9(a)), the methods are indistinguishable (ignoring *weighted position*—see caption). At a lead time of 2 d 3 h (figure 9(b)), the high resolution forecast has the edge during this period and the ensemble mean methods are indistinguishable from each other. The high resolution forecast almost always out-performs the control run; this is unsurprising given the difference in resolution.

The key result is the performance of the *weighted position* trading strategy. This outperforms all the *fixed position* methods at all lead times. The best example of the performance of *weighted position* is in figure 8. Inspecting figure 8(b) reveals why the method works so well. It rarely performs worse than the fixed-position methods; its daily profit is almost always greater than or equal to theirs. On the other hand, there are many occasions when it performs significantly better than the fixed-position methods which accumulate to make a large profit. Spikes of high profit from *weighted position* tend to line up with spikes of moderate profit from the fixed-position methods, especially where they agree closely (indicated by the shaded region being narrow). These cases where the forecasts agree closely will tend to be cases where the atmosphere is more predictable. The ensemble is successfully identifying the more predictable atmosphere by exhibiting small spread, resulting in large trades and therefore profits. It does this often enough that *weighted position* generates significantly more profit overall than *fixed-position* methods. Similar small scale features can also be identified in figure 9.

This finding has one caveat: at short lead times, the under-dispersion of the ensembles means that the mean position is significantly greater than 100 MWh. This results in the profit being almost twice that from using even a perfect forecast (the final profit of *perfect forecast* should be considered the maximum obtainable profit over any given time period). The results at these lead times should therefore be considered spurious.

Finally, the *best member* forecast produces extremely good results, largely independent of lead time. This demonstrates that the ensemble always contains an extremely accurate forecast, and there is therefore room for improvement on current methods.

4. Conclusions

NWP forecasts are already used, daily, in energy trading markets. Whilst the high resolution runs and the ensemble means are used more than the ensemble members, the latter are regarded as giving a measure of forecast uncertainty. Indeed, advertisements for

software platforms use the inclusion of ensembles as a selling point. The adoption of methodologies such as those used in this study depend not only on their success, but also on the simplicity and efficiency of implementation.

In this study we have found a successful method to convert an ensemble wind speed forecast into an ensemble forecast of aggregated wind power across Germany. We demonstrated that all our numerical forecasting methods exhibit positive skill, compared to the climatology, out to a lead time of seven days. We further compared the forecast skill of multiple methods and found that ensemble methods exhibit significantly more skill than traditional, deterministic, methods for medium range forecasts, particularly for lead times exceeding 3 d. We noted the importance of taking into account the nonlinearity of the conversion of wind speed into power, but also found that knowing the precise locations of wind farms is not necessary to produce good medium-range wind power forecasts. An attempt to produce a better point forecast from the ensemble than the ensemble mean was unsuccessful, but a ‘forecast’ generated *a posteriori*, by taking the most accurate ensemble member on each occasion, demonstrated that there is significant scope for improvement on current methods using information currently available in the ensembles.

We also implemented a simple trading model and demonstrated the value of using the probabilistic information in the ensemble to inform decision-making. A trading strategy which weights the size of the trade based on the ensemble spread was found to be significantly more profitable than fixed-trade strategies based on point forecasts alone, with the most reliable results obtained at lead times exceeding 3 days. Upon inspection, it was found that this is because the method identifies occasions when wind power is predictable and makes an accordingly higher-value trade to maximise profits. This demonstrates that ensemble spread is an essential piece of information contained within the ensemble, and can be used to profitable effect by energy market participants.

There has been some suggestion that the performance of methods based on weighting ensemble members could be improved by adjusting those weightings according to weather regime, and one of the authors is working on this idea. Other future work might consider the second-order effects such as wind direction and air density variation which we neglected in the power conversion process.

5. Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The wind power feed-in data used to obtain the wind power generation observations are freely available from the German transmission system

operators TenneT TSO GmbH, 50Hertz Transmission GmbH, Amprion GmbH and TransnetBW GmbH.

Appendix. Ensemble RMSE-spread relation

This derivation uses results from [29]. Equations (A2) and (B1) from that paper are reproduced here as (3) and (4), respectively, using the notation defined therein.

$$\overline{\langle \|X_e - \langle X_e \rangle\|^2 \rangle} = \overline{\langle \|X_T - \langle X_e \rangle\|^2 \rangle} \quad (3)$$

$$\overline{\langle \|X_e - X_T\|^2 \rangle} = 2 \overline{\langle \|X_T - \langle X_e \rangle\|^2 \rangle}. \quad (4)$$

The quantity on the left of (3) can be identified with the mean squared spread of the ensemble, while the left hand side of (4) is the mean squared error of an ensemble member. Substitution of (3) into (4) gives

$$\overline{\langle \|X_e - X_T\|^2 \rangle} = 2 \overline{\langle \|X_e - \langle X_e \rangle\|^2 \rangle}. \quad (5)$$

Taking the square root gives the desired relation; $\text{RMSE} = \sqrt{2} s$.

References

- [1] Federal Ministry for Economic Affairs and Energy 2019 Time Series for the Development of Renewable Energies in Germany 1990–2018 *Government Report* 1990–2018 Germany
- [2] Georgilakis P S 2008 Technical challenges associated with the integration of wind power into power systems *Renew. Sustain. Energy Rev.* **12** 852–63
- [3] Fabbri A, San Román T, Abbad J and Quezada V H M 2005 Assessment of the cost associated with wind generation prediction errors in a liberalised energy market *IEEE Trans. Power Syst.* **20** 1440–6
- [4] Foley A M, Leahy P G, Marvuglia A and McKeogh E J 2012 Current methods and advances in forecasting of wind power generation *Renew. Energy* **37** 1–8
- [5] Lau A and McSharry P 2010 Approaches for multi-step density forecasts with application to aggregated wind power *Ann. Appl. Stat.* **4** 1311–41
- [6] Chen P, Pederson T, Bak-Jensen B and Chen Z 2010 ARIMA-based time series model of stochastic wind power generation *IEEE Trans. Power Syst.* **25** 667–76
- [7] Clark R T, Bett P E, Thornton H E and Scaife A A 2017 Skilful seasonal predictions for the European energy market *Environ. Res. Lett.* **12** 024002
- [8] Ohba M, Kadokura S and Nohara D 2018 Medium-range probabilistic forecasts of wind power generation and ramps in Japan based on a hybrid ensemble *Atmosphere* **9** 423
- [9] Lee D Y, Doblas-Reyes F J, Torralba V and Reviriego G 2019 Multi-model seasonal forecasts for the wind energy sector *Clim. Dyn.* **53** 2715–29
- [10] Taylor J W, McSharry P E and Buizza R 2009 Wind power density forecasting using ensemble predictions and time series models *IEEE Trans. Energy Convers.* **24** 775–82
- [11] Palmer T N 2018 The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years *Q. J. R. Meteorol. Soc.* **145** 12–24
- [12] Tennekes H, Baede A P M and Opsteegh J D 1987 Forecasting forecast skill *Proceedings of the Workshop on Predictability in the Medium and Extended Range* (Reading: ECMWF) pp 277–302 (www.ecmwf.int/node/12652)
- [13] Molteni F, Buizza R, Palmer T N and Petroliagis T 1996 The ECMWF ensemble prediction system: methodology and validation *Q. J. R. Meteorol. Soc.* **122** 73–199
- [14] Traiteur J J, Callicut D J, Smith M and Roy S B 2012 A short-term ensemble wind speed forecasting system for wind power applications *J. Appl. Meteorol. Climatol.* **51** 1763–74
- [15] Raftery A E, Gneiting T, Balabdaoui F and Polakowski M 2005 Using Bayesian model averaging to calibrate forecast ensembles *Mon. Weather Rev.* **133** 1155–74
- [16] Pinson P, Chevallier C and Kariniotakis G 2007 Trading wind generation from short term probabilistic forecasts of wind power *IEEE Trans. Power Syst.* **22** 1148–56
- [17] Owens R G and Hewson T D 2018 *ECMWF Forecast User Guide* (Reading: ECMWF) (<https://doi.org/10.21957/m1cs7h>)
- [18] TenneT TSO GmbH. Actual and forecast wind energy feed-in (<https://tennet.eu/electricity-market/transparency-pp/transparency-germany/network-figures/actual-and-forecast-wind-energy-feed-in/>) (Accessed: 30 November 2018)
- [19] 50Hertz Transmission GmbH. Wind power, archived wind power (<https://50hertz.com/en/Transparency/GridData/Windpower>) (Accessed: 30 November 2018)
- [20] Amprion GmbH. Wind feed in (<https://amprion.net/Grid-Data/Wind-Feed-In/>) (Accessed: 30 November 2018)
- [21] TransnetBW GmbH. Wind feed in (<https://transnetbw.com/en/transparency/market-data/key-figures>) (Accessed: 30 November 2018)
- [22] Sanchez I 2006 Short-term prediction of wind energy production *Int. J. Forecast.* **22** 43–56
- [23] Cannon D J, Brayshaw D J, Methven J, Coker P J and Lenaghan D 2015 Using reanalysis data to quantify extreme wind power generation statistics: a 33 year case study in Great Britain *Renew. Energy* **75** 767–78
- [24] Vestas wind systems A/S. General Specification V90-3MW, 2004 (www.gov.pe.ca/photos/sites/envengfor/file/950010R1_V90-GeneralSpecification.pdf)
- [25] Talagrand O, Vautard R and Strauss B 1997 Evaluation of probabilistic prediction systems *Proceedings, ECMWF Workshop on Predictability* (Reading: ECMWF) pp 1–25 (www.ecmwf.int/node/12555)
- [26] Wilks D S 2011 *Statistical Methods in the Atmospheric Sciences (International Geophysics Series Vol 100)* 3rd edn (London: Academic)
- [27] Finney I 2016 Measuring the value of ensemble forecasts to renewable energy: so much more than the ensemble mean *Using ECMWF's Forecasts Workshop* (Reading: ECMWF) (www.ecmwf.int/node/16433)
- [28] Leith C E 1974 Theoretical skill of Monte Carlo forecasts *Mon. Weather Rev.* **102** 409–18
- [29] Palmer T N, Buizza R, Hagedorn R, Lawrence A, Leutbecher M and Smith L 2005/06 Ensemble prediction: a pedagogical perspective *ECMWF Newsletter* **126** 10–7