# *Determining the bounds of skilful forecast range for probabilistic prediction of system-wide wind power generation*

Meteorologische Zeitschrift, PrePub DOI 10.1127/metz/2016/0751




# Determining the bounds of skilful forecast range for probabilistic prediction of system-wide wind power generation

Dirk Cannon[1*], David Brayshaw[1,2], John Methven[1] and Daniel Drew[1]

[1]Department of Meteorology, University of Reading, UK
[2]National Centre for Atmospheric Science, University of Reading, UK

(Manuscript received October 31, 2015; in revised form March 25, 2016; accepted June 6, 2016)

**Abstract**

State-of-the-art wind power forecasts beyond a few hours ahead rely on global numerical weather prediction models to forecast the future large-scale atmospheric state. Often they provide initial and boundary conditions for nested high resolution simulations. In this paper, both upper and lower bounds on forecast range are identified within which global ensemble forecasts provide skilful information for system-wide wind power applications. An upper bound on forecast range is associated with the limit of predictability, beyond which forecasts have no more skill than predictions based on climatological statistics. A lower bound is defined at the lead time beyond which the *resolved* uncertainty associated with estimating the future large-scale atmospheric state is larger than the *unresolved* uncertainty associated with estimating the system-wide wind power response to a given large-scale state.

The bounds of skilful ensemble forecast range are quantified for three leading global forecast systems. The power system of Great Britain (GB) is used as an example because independent verifying data is available from National Grid. The upper bound defined by forecasts of GB-total wind power generation at a specific point in time is found to be 6–8 days. The lower bound is found to be 1.4–2.4 days. Both bounds depend on the global forecast system and vary seasonally. In addition, forecasts of the probability of an extreme power ramp event were found to possess a shorter limit of predictability (4.5–5.5 days). The upper bound on this forecast range can only be extended by improving the global forecast system (outside the control of most users) or by changing the metric used in the probability forecast. Improved downscaling and microscale modelling of the wind farm response may act to decrease the lower bound. The potential gain from such improvements have diminishing returns beyond the short-range (out to around 2 days).

**Keywords:** Wind Power Forecasting, Global NWP, Ensemble Forecasting, Predictability, Ramping, Skill Scores

## 1 Introduction

Global wind power capacity has risen dramatically from around 8 GW in 1997 to around 370 GW at the end of 2014 (GWEC, 2015). Consequently, wind power forecasts have become an essential tool for energy market participants and system operators in many countries. Accurate forecasts contribute to efficient strategies for market trading and asset maintenance, as well as for efficient unit commitment and system balancing, whereby enough reserve power must be available to maintain security of supply (Giebel et al., 2011).

To operate the energy system efficiently, different types of forecast are required at a range of different lead times. For example, forecasts of power output at a specific time point in the future are widely used at a variety of lead times from minutes ahead to weeks ahead (Morales et al., 2014). Forecasts for extreme events such as large ramps in wind power generation are also used by market participants and transmission system operators (Cutler et al., 2007; Greaves et al., 2009; Ferreira et al., 2010). In addition, weekly-mean wind forecasts can also be used by energy market participants at longer lead times (at 3–4 weeks ahead, Lynch et al., 2014). For each user, optimal decisions must be made by weighing the costs and potential losses associated with taking decisions against the potential returns of acting early, for example by applying cost-loss functions to a probability forecast (Richardson, 2012).

Up to a few hours ahead of real time, statistical methods are used to predict the future wind power output[1]. Beyond this time horizon, forecasts rely heavily on global numerical weather prediction (NWP) models (Giebel et al., 2011). NWP model forecasts are initialised from *analyses*, which represent the observed state of the whole atmosphere represented on the three-dimensional NWP model grid by blending observational data with an earlier forecast using a data assimilation method.

*Corresponding author: Dirk Cannon, University of Reading, Whiteknights, PO Box 217, Reading, Berkshire RG6 6AH, United Kingdom, e-mail: d.j.cannon@reading.ac.uk

[1]In the meteorological community, forecasting up to a few hours ahead is often termed *nowcasting*.

A major source of uncertainty in NWP model forecasts is associated with the chaotic behaviour of the atmosphere, whereby the state at a future time is sensitive to small changes at the start of the forecast (Lorenz, 1963). As the initial state is uncertain, this uncertainty grows with increasing lead time, and it is this uncertainty that global forecast centres attempt to capture using forecast ensembles. The ensemble members are constructed by adding perturbations to the analysis to represent the range of observational uncertainty in the initial atmospheric state. The forecast model is run forward from each of these slightly different initial conditions and the ensemble perturbations are calibrated so that the ensemble spread matches the forecast error (averaged over many start dates) at specific lead times of interest; for example at 2 days ahead for the European Centre for Medium Range Weather Forecasts (ECMWF) forecast system. The spatial structure of the perturbations also affects the rate of spread and the locations with the greatest spread. The techniques used to construct the perturbations differ between global forecasting centres, and so the characteristics of their spatial structure differs along with their evolution and the average rate of increase in spread. In addition, parameters describing physical processes in the NWP model can be perturbed stochastically during the forecast calculations, which acts to increase further the differences between ensemble members and the rate of increase in the ensemble spread.

For a global ensemble forecast at lead time $l$, initialised at time $t_i$ and valid at time $t_v = t_i + l$, the spread of the ensemble relative to the analysis at the same valid time ($t_v$) is herein referred to as the *resolved uncertainty*. As analyses are used to initialise the global forecasts, the *resolved uncertainty* is small when $l = 0$ (reflecting the ensemble perturbations only) and grows with increasing lead time. Eventually the ensemble spread must become so large that there is no advantage in using the forecast over using climatological statistics. For example, in Sections 3.1-3.2 the quantity being forecast (the *predictand*) is the system-aggregated capacity factor, $c(t_v)$, which is defined here as the total wind power generation at valid time $t_v$ expressed as a percentage of the total installed wind capacity[2]. An ensemble forecast can be used to generate a histogram where the range of $c$ is divided into bins and the number of forecast members within each bin ($c \rightarrow c + \delta c$) is counted. This is often interpreted as a forecast of the PDF (probability density function) of the predictand. For example, the probability that $c$ will lie between 25–30 % at valid time $t_v$ can be inferred from the number of ensemble members predicting $c$ within that range. Typically the histogram becomes broader and flatter with increasing lead time as the uncertainty increases. At long lead times the ensemble PDF usually approaches the *climatological* PDF obtained from the long-term statistics of a free-running model. If the model is sufficiently well calibrated this will also match the climatological PDF obtained from the long term statistics of the observed $c$. The characteristic lead time at which the skill of ensemble forecasts is on average no better than a climatological prediction is referred to as the limit to predictability of the first kind (Lorenz, 1975). Here, we refer to this limit as the *upper bound*, $t_U$, of the useful forecast range.

Wind estimates based on global NWP analyses can differ substantially from observed values (Pinson and Hagedorn, 2012). Here, the difference between the analysis estimate of GB-total wind power generation and the observed power output will be referred to as the *unresolved uncertainty* since it relates to processes and phenomena that are not explicitly resolved by the global NWP model. Contributions to the *unresolved uncertainty* include (i) the uncertainty in local wind speed features that are not resolved by the global model, (ii) the uncertainty in the response of the wind farm to those winds, and (iii) the uncertainty stemming from the operational management of the power system, including instructions to curtail power generation, unreported maintenance and metering errors.

To analyse a large number of forecasts without incurring excessive computational costs, here we adopt a relatively simple method to produce wind power generation estimates from the global NWP model winds (described in Section 2). However, in operational forecasting systems a number of methods are employed to further reduce the *unresolved uncertainty* as defined above. One approach is to dynamically downscale the global NWP data using a regional NWP model to provide weather information at higher spatial and temporal resolution. Dynamical downscaling is particularly useful in regions of complex terrain where the topography is poorly resolved by global NWP models (Reid and Turner, 2001; Mass et al., 2002; Jiménez et al., 2010). Information from NWP models is also often used as input to physical-statistical *microscale* models that estimate the power output for individual wind farms. Such models vary in complexity, however a common approach (which is adopted in this paper) is to utilise a *power curve*, which is a simple transformation relating the power generated to the ambient wind speed. Power curves can be derived from knowledge of the turbine design (Carrillo et al., 2013) or by empirically deriving statistical relationships using measured data (Giebel et al., 2011). Local forecasts can also be re-calibrated and improved using statistical techniques such as model output statistics (Glahn and Lowry, 1972). For some forecast users, such as transmission system operators, the local wind farm forecasts are spatially aggregated over a region of interest. This region can be system-wide (as in this paper), or defined within a zone subjected to constrained transmission at its boundaries.

The *total uncertainty* in wind power forecasts is thus made up of both *resolved* and *unresolved* contributions. The *resolved uncertainty* must increase with lead time, as a result of chaos, and at some lead time ($t_L$), come to dominate the *total uncertainty*. In order to estimate $t_L$ it is essential to consider a problem where the predictand

[2]The capacity factor is also commonly referred to as the *load factor*.

**Table 1:** A summary of the data sources used in this project, where N is one quarter of the number of grid points around the equator, and abbreviations are defined in the text. Note that the forecast systems in TIGGE are upgraded periodically, and only the latest (maximum) horizontal grid resolution is shown here. For more detailed information, see Bougeault et al. (2010) and Table 1 of Gray et al. (2014). All TIGGE data is interpolated onto a common 0.5 ° by 0.5 ° horizontal grid prior to archiving. For more detailed information about the ERA-Interim data, see Dee et al. (2011).

| Name or Abbrev. | Data Type | Horizontal Discretisation | Source |
|---|---|---|---|
| GB Wind Power | Measured | National | National Grid |
| ECMWF | Global Forecasts and Analyses | Max. N320 (≈ 16 km) | TIGGE |
| UKMO | Global Forecasts and Analyses | Max. N216 (≈ 60 km) | TIGGE |
| NCEP | Global Forecasts and Analyses | Max. N128 (≈ 80 km) | TIGGE |
| ERA-Interim | Reanalyses | N128 (≈ 80 km) | ECMWF |

is well observed by measurements that are completely independent of atmospheric observations or models. In this case, the measurement is the wind power generation coming onto the Great Britain (GB) electricity transmission system. Improved understanding and forecasting of the wind resource is particularly important in GB due to the large and increasing market penetration of wind power and its relative isolation from other European systems (transnational electricity flows are small compared to overall GB demand). It is shown that $t_L$ varies for forecasts from different global NWP models and we use it to define the *lower bound* of useful ensemble forecast range.

Understanding the relative contribution of *resolved* and *unresolved* sources of uncertainty provides forecast users and developers with guidance as to the maximum potential value of improvements to their forecast systems for sytem-wide wind power applications. For example, improvements to the global NWP forecast system may extend the upper bound $t_U$. However, such improvements may also reduce the *resolved uncertainty* at shorter lead times, increasing the lower bound $t_L$ if the *unresolved uncertainty* remained unchanged. The useful forecast window may be increased by reducing $t_L$ through improvements to the downscaling and physical-statistical modelling of the wind farm response, thereby reducing the *unresolved uncertainty*.

In Section 2, the process of converting NWP model wind speeds to wind power generation estimates is described, after which results are presented in Section 3 and conclusions in Section 4. The results in Section 3 are presented in three parts. In Section 3.1, the upper bound of useful forecast range, $t_U$, is characterised using ensemble forecasts of GB wind power derived from output of three leading global forecast centres. In Section 3.2, the lower bound, $t_L$, is quantified, below which the *unresolved uncertainty* in the predictand (GB wind power generation) is larger than the *resolved uncertainty*. In Section 3.3, global ensembles are used to forecast the probability that wind power swings (ramps) exceeding a certain threshold will occur within a pre-determined time-window. The limit to predictive skill in the forecast of this probability, $t_P$, is compared with the $t_U$ and $t_L$ estimated for GB wind power to assess the range of lead times for which such probability forecasts can be useful for risk-based decision making.

## 2 Method

In this paper, the national wind power generation derived from global meteorological forecasts and analyses are compared to system-aggregated wind power observations from the Great Britain (GB) power system, provided by National Grid. The analyses and forecasts are obtained through the data portal of the THORPEX Interactive Grand Global Ensemble (TIGGE)[3]. The TIGGE dataset was created as an activity of the World Meteorological Organisation's THORPEX Programme. Ensemble forecast output from 10 global operational weather forecasting centres has been archived in one place to enable utilisation of forecast data in atmospheric research. See Bougeault et al. (2010) for a description of the purpose and initial use of TIGGE data since its start in December 2006, and Swinbank et al. (2016) for some highlights of research using TIGGE over its first decade.

For the global NWP forecasts and analyses, wind speeds are archived in the TIGGE database at 10 m above the surface and at various vertical levels corresponding to constant pressure surfaces. Wind speeds at 10 m above the surface and at the two lowest pressure levels (925 and 850 hPa) from three internationally recognised global forecast providers are used: (i) the European Centre for Medium Range Weather Forecasts (ECMWF), (ii) the UK Met Office (UKMO), and (iii) the National Centers for Earth Prediction (NCEP), from the USA. Whilst the forecasts and analyses were originally produced at different resolutions (Table 1), they were pre-interpolated (during archival) onto a regular 0.5 ° by 0.5 ° horizontal grid in order to more easily facilitate multi-model comparisons (e.g., Johnson and Swinbank, 2009; Hagedorn et al., 2012).

The forecast and analysis wind speeds were converted to GB-aggregated wind power generation following the general approach of Cannon et al. (2015). The steps are: (i) bi-linearly interpolate the wind speeds horizontally from the global NWP model grid to the location of each wind farm (Figure 1), (ii) assume a logarithmic change in wind speed with height to interpolate the

[3] http://apps.ecmwf.int/datasets, accessed 28 October 2015).

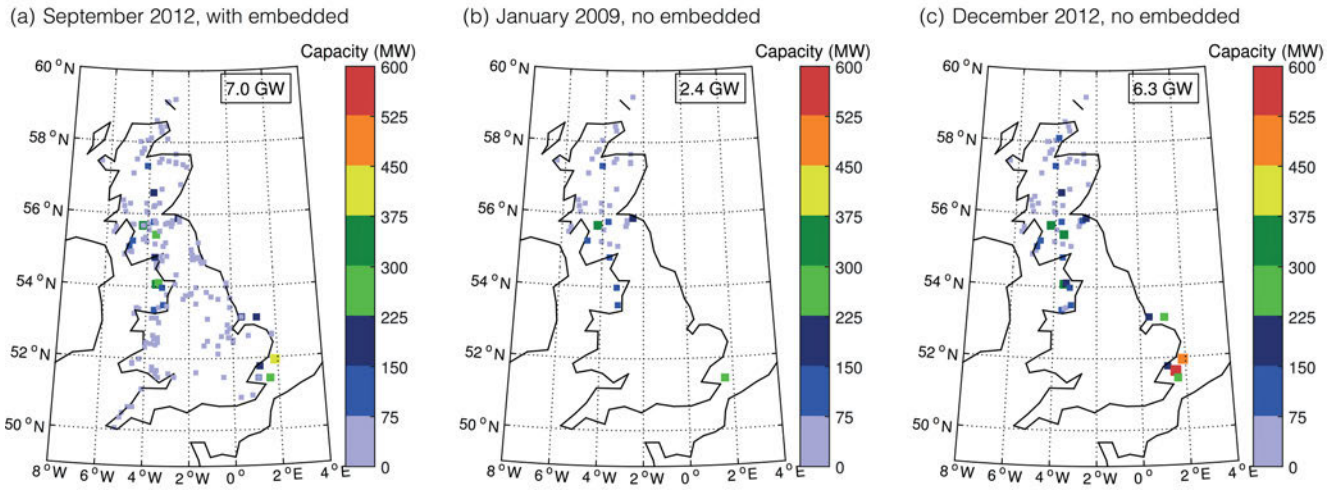


**Figure 1:** (a) The distribution and capacity of wind farms for the Great Britain (GB) power system in circa September 2012 (as in Cannon et al., 2015). This is the fixed distribution assumed in Section 3.1, and includes many wind farms that are embedded within local electricity distribution networks (including all onshore wind farms in England and Wales). In Section 3.2, a time-varying distribution is used to provide a more robust comparison with measured wind power data from National Grid. Here, the distribution evolves from (b) 01 January 2009 to (c) 31 December 2012, and wind farms embedded within local distribution networks are not included. The GB-total capacity for each distribution is shown in the top right.

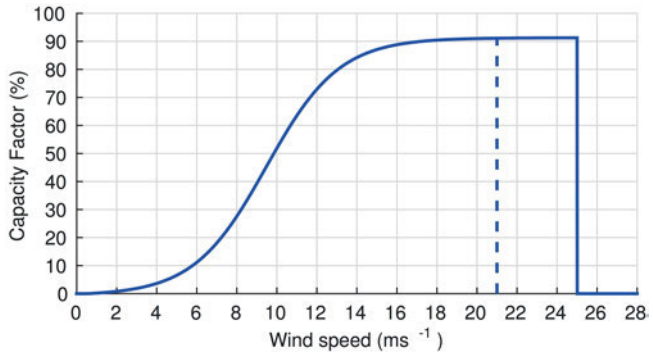


**Figure 2:** The assumed dependence of wind farm power output on hub-height wind speed (Cannon et al., 2015). High wind speed cut-out occurs above 25 $ms^{-1}$ after which wind farms come back online at 21 $ms^{-1}$ (dashed).

winds from the available vertical model levels to a representative turbine hub height, (iii) infer the wind farm power output from the hub-height wind speed assuming a simple deterministic relationship, or *power curve* (Figure 2), and (iv) aggregate over all the wind farms in the distribution. Here, the wind speeds at 10 m above the surface and (for almost all cases) at the 925 hPa pressure level were used to vertically interpolate to hub height. On the few occasions when the 925 hPa level dropped below 20 m above the surface the 850 hPa level winds were used instead of the 925 hPa level winds (this occurs for less than 0.1 % of all forecasts). The power curve was fitted so as to remove the mean bias with respect to measured wind power generation from 2012.

The ability of global meteorological analysis data to estimate GB-aggregated wind power amount and its variability was investigated in detail by Cannon et al. (2015) using the MERRA dataset (Rienecker et al., 2011)[4]. The GB-aggregated wind power estimates were highly correlated to the observed GB wind power generation in 2012, producing a correlation coefficient of 0.96. The variability was well reproduced on time scales of around 6 hours or more (correlation coefficients of 0.77, 0.86 and 0.93 for time scales of 3, 6 and 12 hours respectively).

The global NWP analyses are available every 12 hours, when the forecasts were initialised at 00 and 12 UTC each day. However, the data from each forecast is available at 6 hour time intervals up to 10 days lead time, at 00, 06, 12 and 18 UTC each day. Therefore, for the assessment of ramp forecasts (Section 3.3), the unperturbed (*control*) forecasts at 6 hours ahead are used as a proxy for the analysis at 06 and 18 UTC (initialised from the analyses at 00 and 12 UTC respectively). This approach allows for the assessment of each forecast relative to its own verifying analysis. The alternative approach of using different analysis datasets that contain higher frequency data such as MERRA (Rienecker et al., 2011) or ERA-Interim (Dee et al., 2011), was found to introduce additional biases which favoured some forecast systems more than others.

The distribution of GB wind farms has changed somewhat over the period studied (Figure 1), with a greater proportion of wind capacity now located in the south and offshore, in particular in the Thames Estuary (at approximately 1.5 ° E, 51.5 ° N). In Sections 3.1 and 3.3, only the *resolved* uncertainty (*forecasts vs. analyses*) is considered, and so the fixed wind farm distribution of Cannon et al. (2015), from circa September 2012, is used (Figure 1(a)). This distribution in-

[4]A GB-aggregated wind power time series based on this MERRA data, along with the underlying Matlab model is freely available at http://www.met.reading.ac.uk/~energymet/data/Cannon2015/.

cludes many wind farms that are embedded within the local distribution networks (including all onshore wind farms in England and Wales) and so closely represents the true distribution of the GB wind fleet at that time. In Section 3.2 however, the GB power generation derived from global NWP model data is compared directly to the measured GB wind power data and so the embedded wind farms are excluded as they do not contribute to the wind power measured through the national transmission system (Figure 1(b,c)). The addition of new wind farms, particularly offshore, can significantly alter the characteristics of the wind resource (Drew et al., 2015). Therefore, to further aid the comparison in Section 3.2 the distribution of wind farms is allowed to evolve, with the start date of new wind farms estimated by National Grid from the measured data itself, and cross-checked against publicly available information from company websites.

# 3 Results

In Sections 3.1 and 3.2, the upper and lower bounds on the useful forecast range for ensemble forecasts of Great Britain (GB) capacity factor, $c$ ($t_U$ and $t_L$, respectively), are quantified based on global weather forecasts from three leading global forecast centres (ECMWF, UKMO and NCEP). The metrics used focus on the closeness of the ensemble of forecast $c$ values to the observed $c$ as a function of forecast lead time. In Section 3.3, forecasts for the probability of extreme wind power ramp events are investigated. The limit of predictability associated with these probability forecasts, $t_P$, is compared with the upper bound on forecast range associated with predicting $c$ values.

## 3.1 Upper bound of useful forecast range ($t_U$)

The lead time beyond which wind power forecasts perform no better, on average, than climatological predictions is known to be sensitive to the quantity being forecast. For example, Frame et al. (2011) showed that forecasts of wind speed averaged over the north Atlantic region can retain skill out to beyond 10–15 days relative to climatology, whilst Lynch et al. (2014) found that forecasts of weekly-mean wind speed over the UK retain skill to at least 3 weeks ahead.

Here, the GB wind power predicted by global NWP forecast systems is compared with that based on the global NWP analyses. This quantifies only the *resolved uncertainty*, as defined in Section 1. A long history of forecasts from the TIGGE archive (7 years, from 2007 to 2013 inclusive) are used, over which time the GB wind farm distribution has changed dramatically. As discussed in Section 2, a fixed wind farm distribution from September 2012 is used for this calculation which includes many wind farms embedded within local distribution networks, since this better reflects the contemporary distribution of wind farms across the country. This approach allows for the influence of atmospheric variability and predictability on power generation to be partitioned from the influence of the changing power network.

The accuracy of an ensemble forecast of a single metric (GB capacity factor, $c$) relative to the analyses can be assessed for a single point in time using the continuous ranked probability score (Hersbach, 2000),

$$\mathrm{CRPS}_{\mathrm{forecast}}(l) = \frac{1}{N}\sum_i \int_{c=0\,\%}^{c=100\,\%} (F_i(l) - A_i)^2\, dc, \tag{3.1}$$

where $F_i(l)$ and $A_i$ refer to the cumulative density functions (CDFs) for values of $c$ obtained from the forecast ensemble and analysis (for valid time point $i$ and lead time $l$). As the analysis only has one $c$ value on each time point, $a_i$, its CDF consists of a step function with the transition from zero to one at the point $c = a_i$. The CRPS is a commonly used scoring rule that quantifies the proximity of the ensemble forecast members to the analysis value. It is strictly proper in that it produces a minimum score (zero) only when all ensemble forecast members agree exactly with the analysis value. Forecast members far from the analysis are penalised according to the square of their distance.

To assess the added value of the forecast ensembles relative to a climatology-based statistical prediction, we use the cumulative ranked probability skill score,

$$\mathrm{CRPSS} = 1 - \frac{\mathrm{CRPS}_{\mathrm{forecast}}}{\mathrm{CRPS}_{\mathrm{climatology}}}, \tag{3.2}$$

where $\mathrm{CRPS}_{\mathrm{climatology}}$ is defined as in (3.1) but with the substitution of a CDF derived from a climatology of statistics from the analyses in place of the CDF from the forecast ensemble $F_i$. The climatological CDFs were constructed using the analysis-derived capacity factor values within a moving 7-day window centred on the time point in question. Only the values from the same time of day (00 or 12 UTC) were used to compile the CDFs, thus allowing for diurnal variation in the climatological statistics. For each time point there are thus 7 values of $c$ (one per day in the 7-day window) for each of the 7 years in the TIGGE record, giving at most 49 values contributing to the climatology CDF for each date and time[5]. For example, the climatological ensemble at 12 UTC on 31 August of each year is made up of the 12 UTC values from 28 August to 03 September in each of the years from 2007–2013. A climatological distribution using a moving 30 day window was investigated but found to produce less accurate predictions than the 7 day window used here. A perfect ensemble forecast results in CRPSS = 1, whereas CRPSS = 0 indicates a forecast that offers no more information, on average, than a climatological prediction. A negative score implies that

[5]This is comparable to the number of ensemble members in the ECMWF forecast system (51), and considerably more than the UKMO (24) or NCEP (21) forecast systems.

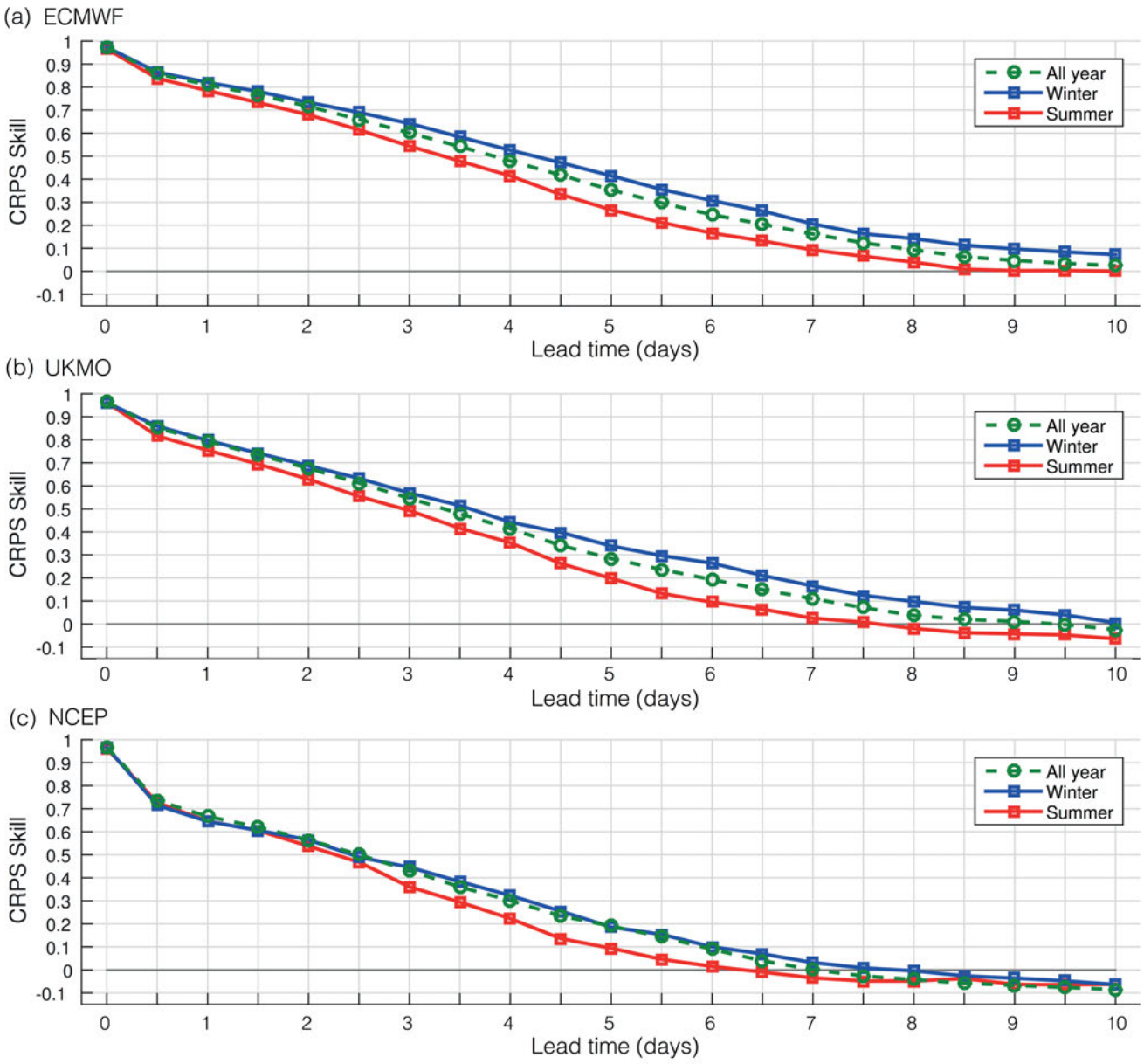


**Figure 3:** The continuous ranked probability skill score (CRPSS), for (a) ECMWF, (b) UKMO and (c) NCEP forecasts of GB-aggregated wind power generation at a future point in time. The skill during summer months (June, July and August; red) and winter months (December, January and February; blue) are shown as well as the all-year mean skill (green).

the ensemble forecast is less accurate on average than the climatological prediction based on the analyses.

Figure 3(a) shows CRPSS as a function of forecast lead time for the ECMWF forecast system. At short lead times the skill is slightly less than perfect ($CRPSS = 1$) due to the perturbations added to each ensemble forecast member at initialisation. As the lead time increases, the skill reduces and eventually asymptotes towards zero, indicating that at long lead times the forecast system is no better than a prediction based on climatological statistics. Even though the CRPS values for the ensemble forecasts are greatest in summer, the skill relative to the climatology is considerably lower in summer than in winter. This is due to the lower variability in summer, resulting in much improved accuracy for the climatological predictions which more than offsets the improvement in the forecast. As a result, the forecast horizon beyond which the forecasts provide negligible skill relative to the climatological prediction is seasonally dependent. To account for the skill falling asymptotically towards zero, we use a threshold of CRPSS = 0.1 to define the upper limit of predictability. This is preferable to assuming skill in the forecasts at long lead times when values can remain marginally above zero for a wide range of lead times. Whilst on average the forecast skill drops below CRPSS = 0.1 beyond around 8 days ahead, in summer this limit is only around 7 days ahead whilst in winter it is around 9 days ahead.

Figures 3(b,c) show equivalent plots for the UKMO and NCEP forecast systems, where the reference cli-

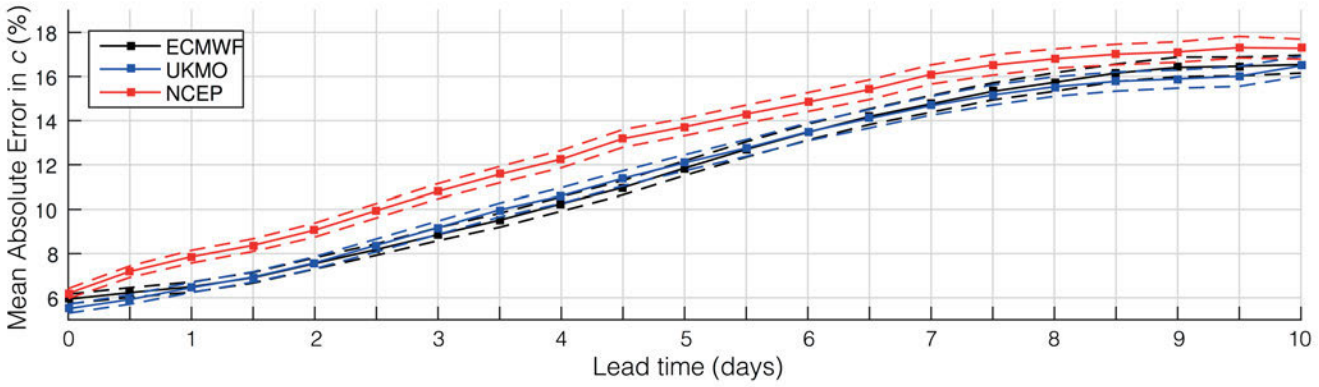


**Figure 4:** The mean absolute error for ensemble mean forecasts of GB-aggregated capacity factor ($c$), relative to National Grid observations (as defined in Eq. 3.3). Dashed lines indicate the 95 % confidence intervals, calculated using a bootstrapping technique.

matological forecast is calculated in the same way as before, but using the UKMO and NCEP analyses respectively. Both forecast systems indicate upper limits of predictability that are shorter than that of the ECMWF system (around 7 days for the UKMO system and around 6 days for the NCEP system). For all three forecast systems, the upper limit of predictability for GB wind power generation shown here is significantly shorter than found in the aforementioned studies into weekly-mean wind speed (3–4 weeks) or mean wind speed over a larger geographical area (at least 10–15 days). The negative skill evident for the NCEP forecast system at long lead times is indicative of a systematic model error relative to its analyses (Figure 3(c)).

## 3.2 Lower bound of useful forecast range ($t_L$)

In this section the total mean absolute error,

$$\mathrm{MAE} = \frac{1}{N}\sum_i \left|\overline{f_i}(l) - o_i\right|, \tag{3.3}$$

is used as an average measure of the uncertainty between the forecasts and observations. In (3.3), $o_i$ represents the observed GB-aggregated capacity factor at occasion $i$, $\overline{f_i}(l)$ represents the ensemble mean forecast capacity factor for the same occasion (at lead time $l$ into the forecast), and $N$ is the number of occasions. Using the ensemble mean forecast tends to yield more accurate forecasts on average than using an individual member (such as the control forecast) and is often used at short lead times when the ensemble spread is relatively small. Note that the CRPS converges to the MAE as the ensemble spread tends to zero. The MAE can also be decomposed into resolved and unresolved components, a property that will be utilised later.

Figure 4 shows the total mean absolute error obtained using three state-of-the-art forecast systems (ECMWF, UKMO and NCEP). At all lead times, the ECMWF and UKMO forecast systems outperform NCEP. Note that the verifying observations are calculated from GB electricity transmission system measurements and are completely independent of the atmospheric measurements upon which the forecasts and analyses are based. The ECMWF and UKMO forecast systems show similar total errors, with the UKMO system performing slightly better at short lead times (up to around 1 day ahead) and at long lead times (beyond around 6 days), whereas the ECMWF forecast system performs slightly better between around 1–6 days ahead. For the ECMWF and UKMO forecast systems, the MAE is around 6 % of total GB capacity at short lead times, increasing gradually to around 9 % after 3 days, 12 % after 5 days and around 16.5 % after 10 days.

The dashed lines in Figure 4 show confidence intervals for each MAE value. They are calculated using a bootstrapping technique to create a distribution of possible values for each lead time. The lower and upper dashed lines represent the 2.5th and 97.5th percentiles of these distributions, meaning that there is a 95 % chance of a random sample (with replacement) of 1000 forecast and observation pairs with an MAE falling within that range. Given the analysis includes 2 forecasts per day for 4 years (over 2900 occasions per lead time), the MAE and the component values are tightly constrained. Sampling variability is due primarily to seasonal variations in the MAE, with the lower threshold corresponding to typical summer values and the upper threshold to winter values.

To investigate the contribution of *resolved* errors to the total MAE, the *resolved* ($\mathrm{MAE_R}$) and *unresolved* ($\mathrm{MAE_U}$) mean absolute errors are defined as

$$\mathrm{MAE_R} = \frac{1}{N}\sum_i \left|\overline{f_i}(l) - a_i\right|, \tag{3.4}$$

$$\mathrm{MAE_U} = \frac{1}{N}\sum_i |a_i - o_i|. \tag{3.5}$$

Figure 5 compares $\mathrm{MAE_R}$, $\mathrm{MAE_U}$, and the total MAE, for each forecast system. As the *unresolved* MAE depends only on the analysis-derived GB power generation and the observed generation, it remains constant with increasing lead time. The *resolved* MAE is almost zero at lead time $l = 0$, and increases with lead time, be-

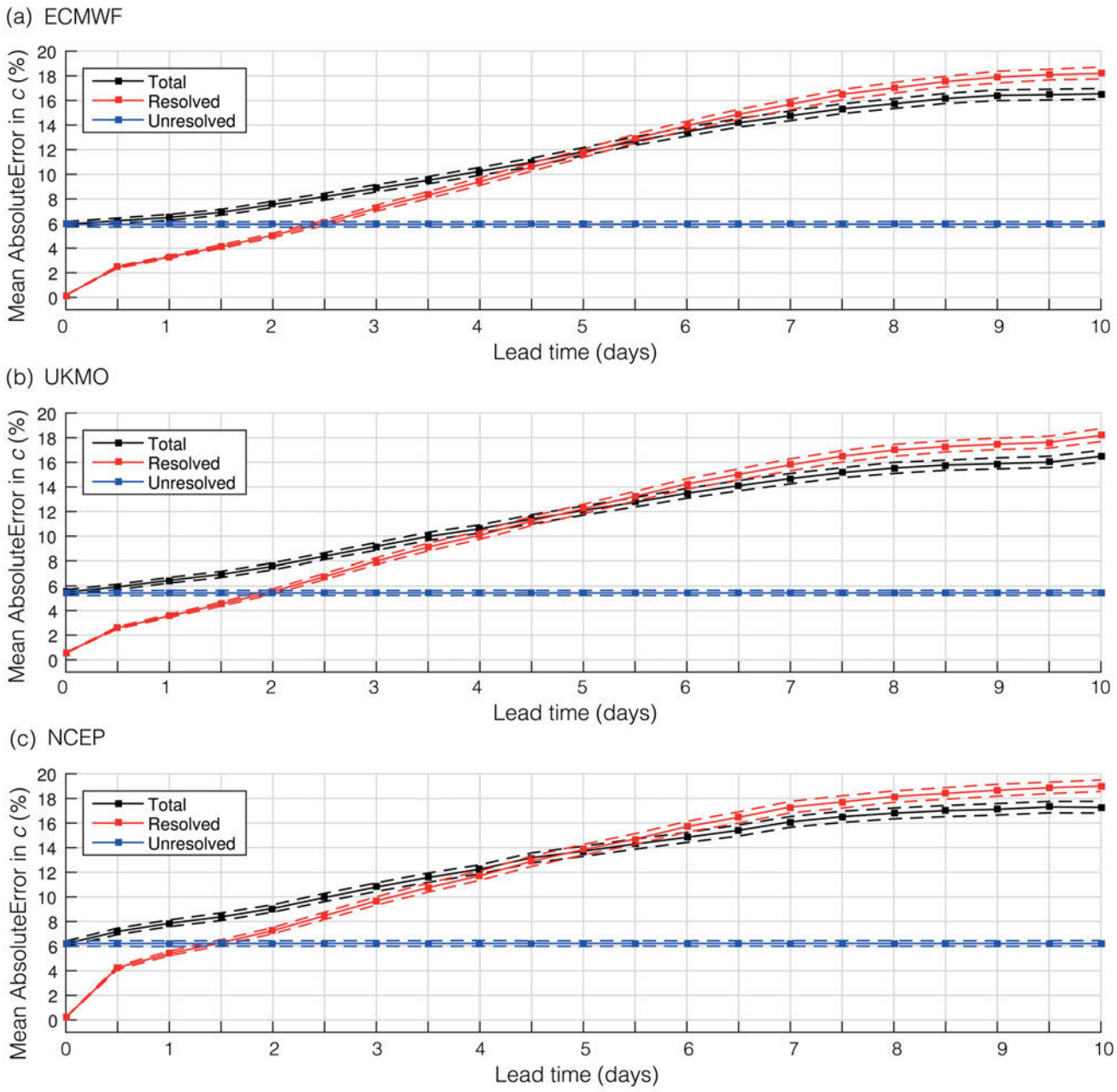


**Figure 5:** The mean absolute error for an ensemble mean forecast of GB-aggregated wind power generation, using the (a) ECMWF, (b) UKMO and (c) NCEP forecast system. The total mean absolute error (MAE in black) is shown alongside the *resolved* ($MAE_R$ in red) and *unresolved* ($MAE_U$ in blue) mean absolute errors. Dashed lines indicate the 95 % confidence intervals, calculated using a bootstrapping technique.

coming larger than the *unresolved* MAE beyond around 1.4–2.4 days ahead depending on the forecast system, and larger than the total MAE beyond around 5 days ahead. Note that the analyses produced by each global centre are used to calculate these statistics, so the magnitude of the *resolved* MAE at the start of forecasts ($l = 0$) reflects only the size of perturbations added to create the varying initial states for each ensemble member. The magnitude of *unresolved* MAE from each centre is very similar, implying that the analyses from all centres produce similar wind speeds over Great Britain, and share common errors relative to the structure of winds incident on wind farm sites.

The fact that $MAE_R$ exceeds the total MAE at long lead times (Figure 5) implies that the distribution of ensemble mean forecasts is closer to that of the observed generation than that of the analyses. This arises because the PDF of capacity factor (from observations, analyses and individual forecasts) is highly skewed due to the positive definite nature of the predictand and the nonlinear conversion of wind speed to capacity factor (Figure 2). Typically in a forecast the ensemble mean CF will exceed the mode, due to the positive skewness, and the PDF of ensemble mean values will be shifted towards higher capacity factor values than the PDF of analyses (a systematic contribution to $MAE_R$). In contrast,

the differences between ensemble mean and observation in each forecast are less systematic. Furthermore, it is found that the PDF of observed CF has slightly higher frequency in the mid-range CF values than analyses (not shown). This leads to smaller differences, on average, between the ensemble mean forecasts and the observations than between the ensemble mean forecasts and the analyses. Note that if the control forecast is used in place of the ensemble mean ($\overline{f_i}(l)$ in equations 3.3 and 3.4) then MAE and $\mathrm{MAE_R}$ agree to within the confidence intervals for 5–10 day lead times, even though both error estimates are greater because for any day's forecast the control could be an outlier while the ensemble mean is likely to be closer to the outcome.

The operation of taking the modulus in the definition of MAE, $\mathrm{MAE_R}$ and $\mathrm{MAE_U}$ means that $\mathrm{MAE} \neq \mathrm{MAE_R} + \mathrm{MAE_U}$. We can alternatively decompose the total MAE into two terms that depend only on either the *resolved* error, $R_i(l) = \overline{f_i}(l) - a_i$, or the *unresolved* error, $U_i = a_i - o_i$. In this manner, the total error can be expressed as the sum of two components:

$$\mathrm{MAE} = \phi_\mathrm{R}(R_i) + \phi_\mathrm{U}(U_i), \tag{3.6}$$

where the *resolved* component, $\phi_\mathrm{R}$, depends only on the *resolved* error, $R_i(l)$, and the *unresolved* component, $\phi_\mathrm{U}$, depends only on the *unresolved* error, $U_i$. This decomposition is achieved by rewriting (3.3) as

$$\mathrm{MAE} = \frac{1}{N}\sum_i |R_i(l) + U_i| = \frac{1}{N}\sum_i \mathrm{AE}_i, \tag{3.7}$$

where $\mathrm{AE}_i = |R_i(l) + U_i|$ represents the absolute (total) error for each occasion $i$ (using the forecast at lead time $l$). The relative impact of $R_i(l)$ and $U_i$ on $\mathrm{AE}_i$ varies depending on one of three possible situations, for which the occasions are divided into the following corresponding sets:

- **Set 1 ($\mathbf{i \in S_1}$)** occurs when $R_i(l)$ and $U_i$ are the same sign, and the errors simply reinforce each other to produce a larger absolute error: $\mathrm{AE}_i = |R_i(l)| + |U_i|$.
- **Set 2 ($\mathbf{i \in S_2}$)** occurs when $R_i(l)$ and $U_i$ are different signs and $|R_i(l)| \geq |U_i|$. In this case, the *unresolved* error ($U_i$) acts to partially or totally compensate for the larger *resolved* error ($R_i(l)$), and so the absolute error is given by the positive difference between them: $\mathrm{AE}_i = |R_i(l)| - |U_i|$.
- **Set 3 ($\mathbf{i \in S_3}$)** is similar to set 2, and occurs when $R_i(l)$ and $U_i$ are different signs but when $|R_i(l)| < |U_i|$. In this case, the *resolved* error ($R_i(l)$) acts to partially compensate for the larger *unresolved* error ($U_i$), and the absolute error is again given by the positive difference between them: $\mathrm{AE}_i = |U_i| - |R_i(l)|$.

Summing over these sets independently, the total mean absolute error (MAE) can be rewritten in terms of the *resolved* and *unresolved* components in (3.6), where

$$\phi_\mathrm{R} = \frac{1}{N}\left[\sum_{i\in S_1}|R_i(l)| + \sum_{i\in S_2}|R_i(l)| - \sum_{i\in S_3}|R_i(l)|\right], \tag{3.8}$$

depends only on $|R_i(l)|$, and

$$\phi_\mathrm{U} = \frac{1}{N}\left[\sum_{i\in S_1}|U_i| - \sum_{i\in S_2}|U_i| + \sum_{i\in S_3}|U_i|\right], \tag{3.9}$$

depends only on $|U_i|$. $\phi_\mathrm{R}$ thus represents the net contribution of all of the *resolved* errors to the overall MAE, and $\phi_\mathrm{U}$ represents the net contribution of all of the *unresolved* errors. In general, the only mathematical constraint on these components is that they sum to MAE (which is positive definite), and so in general if one of them is larger than MAE the other must be negative. However, for a perfectly calibrated (reliable) forecast system, both components should be greater than or equal to zero as set 1 ($i \in S_1$) must account for at least half the total MAE. Note that whilst the *unresolved* errors themselves are independent of the forecast lead time, the number of occurrences in each set ($S_1$, $S_2$ and $S_3$), and therefore $\phi_\mathrm{U}$, does vary with the lead time.

Figure 6(a) shows the proportion of the total MAE attributable to the *resolved* and *unresolved* components ($\phi_\mathrm{R}/\mathrm{MAE}$ and $\phi_\mathrm{U}/\mathrm{MAE}$ respectively), as a function of increasing lead time for the ECMWF forecast system. As in Figure 5, at short lead times the *resolved* errors are small and so the *resolved* error contributes almost nothing to the total MAE. As the lead time increases, the *resolved* component becomes relatively more important as the uncertainty in the large-scale state of the atmosphere increases. In contrast, the contribution of *unresolved* errors decreases with increasing lead time, eventually tending to zero at long lead times. This occurs because, at long lead times, the *resolved* errors are generally larger than the *unresolved* errors and so set 3 occurs increasingly rarely. Therefore, from (3.8–3.9) we see that whilst the *resolved* component ($\phi_\mathrm{R}$) is unequivocally positive, the *unresolved* component ($\phi_\mathrm{U}$) depends on the balance between positive terms (set 1) and negative terms (set 2). At long lead times (when the forecast skill drops to zero) the forecasts become almost random and uncorrelated with the analyses, and so the *resolved* and *unresolved* components are equally likely to act to reinforce each other (set 1) as counteract each other (set 2). Consequently, the positive and negative terms cancel each other out on average and the net impact of small *unresolved* errors is close to zero. This means that small *unresolved* errors added to random large *resolved* errors are as likely to improve the forecast as they are to diminish it.

Figures 6(b,c) show equivalent plots for the UKMO and NCEP forecast systems. Whilst the qualitative behaviour is similar to that in Figure 6(a), the threshold at which the *resolved* uncertainty begins to dominate is markedly different from that of the ECMWF forecast

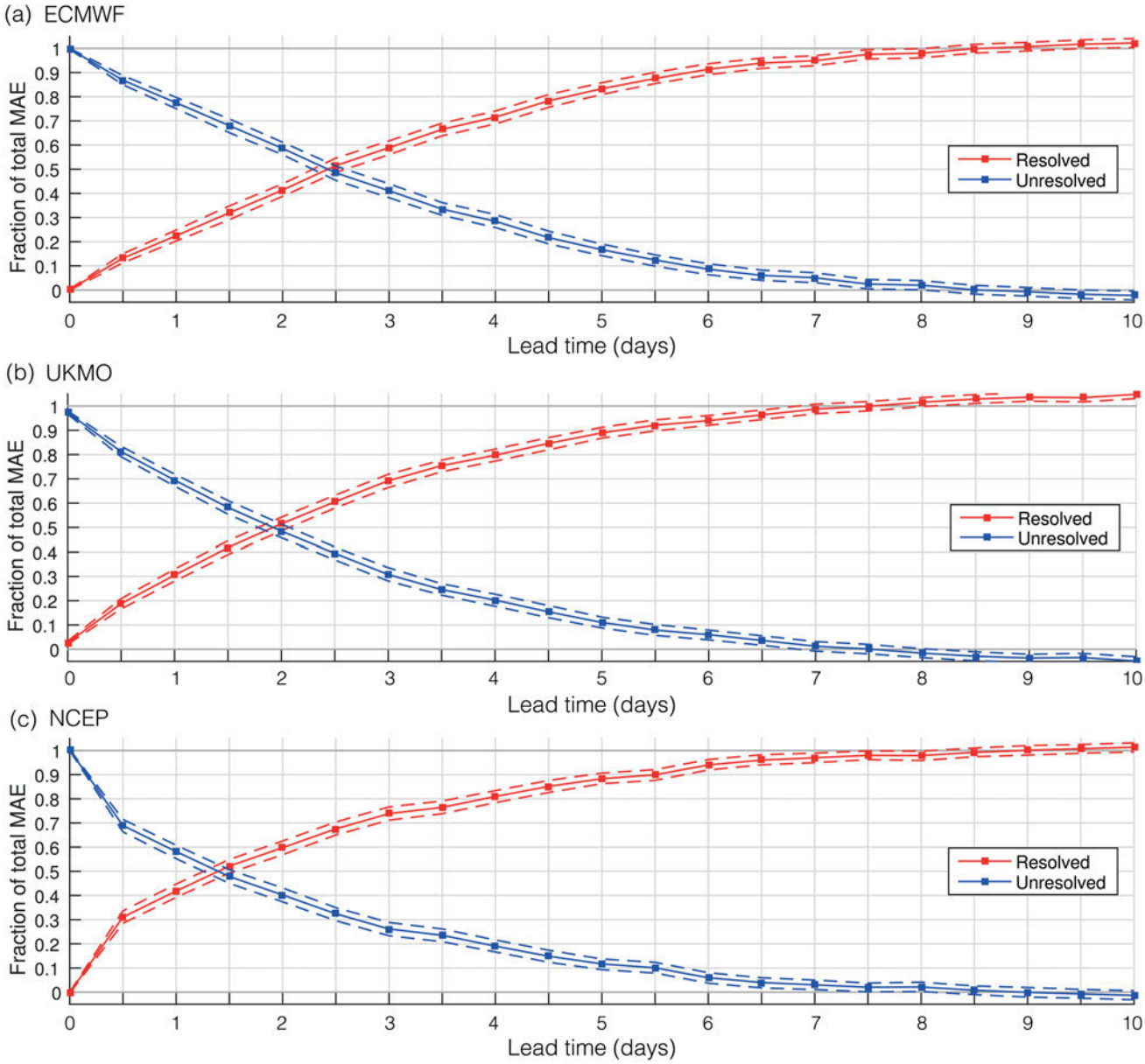


**Figure 6:** The fraction of the total mean absolute error that is attributable to *resolved* ($\phi_R$ in red) and *unresolved* ($\phi_U$ in blue) sources, for the (a) ECMWF, (b) UKMO and (c) NCEP forecast systems. Dashed lines indicate the 95 % confidence intervals, calculated using a bootstrapping technique.

system (around 2.4 days ahead); around 1.9 days ahead for the UKMO forecast system and just 1.4 days ahead for the NCEP forecast system. For UKMO, this difference is due to both a slightly more accurate analysis relative to ECMWF, and a slightly poorer global forecast, resulting in a slightly higher *resolved* MAE and a slightly lower *unresolved* MAE. For NCEP, the short threshold is due mostly to the less accurate global forecast, which leads to a higher *resolved* MAE.

The lower threshold below which the global NWP forecast uncertainty is not the largest contributor to the total uncertainty therefore ranges from around 1.4–2.4 days depending on the forecast system. This suggests that forecast improvements targeted at reducing *unresolved* errors (associated with downscaling and modelling the wind farm response) can significantly improve GB wind power forecasts only at relatively short lead times.

### 3.3 Upper Bound for Probability Forecasts: Extreme Ramps ($t_P$)

At long lead times (days to weeks ahead), transmission system operators are concerned with the likelihood of extreme events such as ramps in wind power generation. Such events occur over short time scales up to a few hours, for example due to local weather phenomena such as convective storms and turbulence. On longer time scales, beyond a few hours, ramp events occur due to the changing synoptic weather patterns, and global

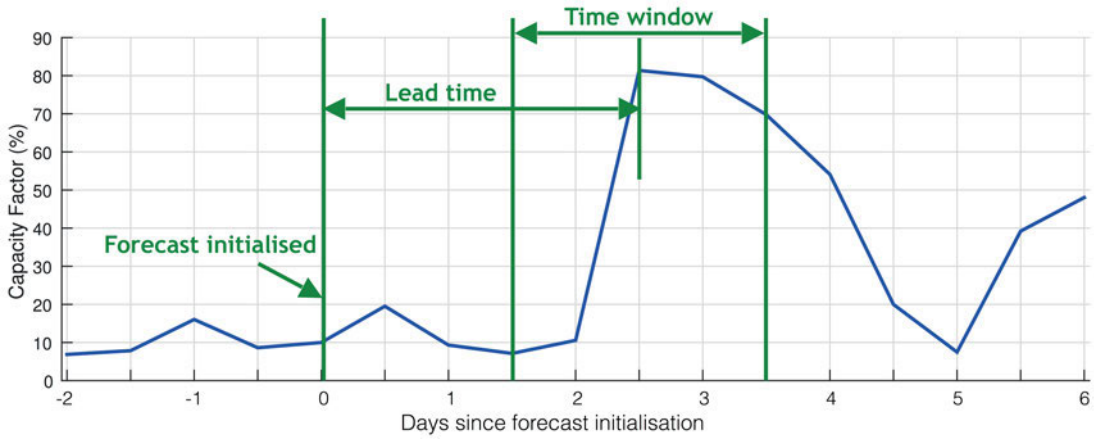


**Figure 7:** An example ramp event. A ramp is said to have occurred if the maximum capacity factor minus the minimum capacity factor within the time window is larger than a defined threshold. The lead time of the ramp is approximated as the time between forecast initialisation and the centre of the time window.

ensemble forecasts enable users to forecast their probability. In this section we assess the skill of ramp forecasts derived from the three global forecast systems and compare the upper limit of predictability to the upper bound on the useful forecast range for the GB capacity factor. As in Section 3.1, only the *resolved uncertainty* is considered here, and so the influence of atmospheric variability is assessed without influence from changes to the power system (by holding the power system fixed). This allows for a longer period of forecasts and analyses to be used which is particularly useful when calculating the probability of infrequent events.

In this paper, a ramp event is said to occur if the maximum absolute change in wind power generation within a given time window exceeds a predefined threshold. The lead time of the ramp is approximated as the time between forecast initialisation and the centre of the time window. An illustrative example is shown in Figure 7, where an increase in capacity factor of around 70 % occurs within a time window of 2 days, at a lead time of 2.5 days. Choosing a ramp threshold of 50 %, a ramp would be said to have occurred here. If however an extremely severe ramp threshold of 80 % were chosen, no ramp event would be said to have occurred in this example. Using this ramp definition, the accuracy with which a forecast system predicts extreme ramp events can be assessed using the Brier Score (Brier, 1950),

$$\mathrm{BS}_{\mathrm{forecast}} = \frac{1}{N}\sum_{i}\left(p_i(l) - q_i\right)^2, \qquad (3.10)$$

where $0 \leq p_i(l) \leq 1$ is the probability of the event occurring in forecast $i$ (at lead time $l$) and $q_i$ is either 0 or 1 (event does or does not occur in the analysis for time-point $i$). As for CRPSS, we define a Brier Skill Score as

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}_{\mathrm{forecast}}}{\mathrm{BS}_{\mathrm{climatology}}}, \qquad (3.11)$$

where the climatological probabilities were constructed using analysis-derived ramp magnitude values within a moving 6-day window. For each time point there are 25 ramp magnitude values for each of the 7 years in the TIGGE record (four per day in the 6 days surrounding the time point, plus one for the time point itself). This gives at most 175 values from which to calculate the probability of the ramp occurring at each time point. As the ramps defined here may occur at any point within a multi-hour or multi-day time window, unlike for the capacity factor climatology in Section 3.1, no diurnal variability is explicitly included in the construction of the climatological probabilities for ramps.

Here, ramps of at least 50 % of GB capacity within a time window of 24 hours are investigated. This corresponds to a swing of 3.5 GW in the (fixed) wind farm distribution (Figure 1(a)). A ramp of this magnitude is large enough to cause significant challenges for transmission system balancing if poorly forecast, and accurate early warnings of such events are extremely helpful in preparing efficient strategies. These ramps occur only infrequently in the reanalysis (7.7 % of the time), and are especially rare in summer, where they occur just 1.0 % of the time[6].

As discussed in Section 2, to address the fact that the analysis data is only available at 00 and 12 UTC whereas the forecasts are available at 00, 06, 12 and 18 UTC, the unperturbed (*control*) forecasts at 6 hours ahead are used to compute values at 06 and 18 UTC (from forecasts initialised at 00 and 12 UTC respectively). Alongside the main results (which will be discussed next), the skill of a 33 year ERA-Interim based climatology relative to this climatology is shown in Figure 8(a–c) for each of the forecast systems. That the ERA-Interim based climatology is only marginally less skilful than the reference climatology used here indicates that the results are not sensitive to the choice of analysis used in the reference climatology.

[6]This means that ramps occur within the subsequent 24 hours of 7.7 % of all time points, and 1.0 % of summer time points.

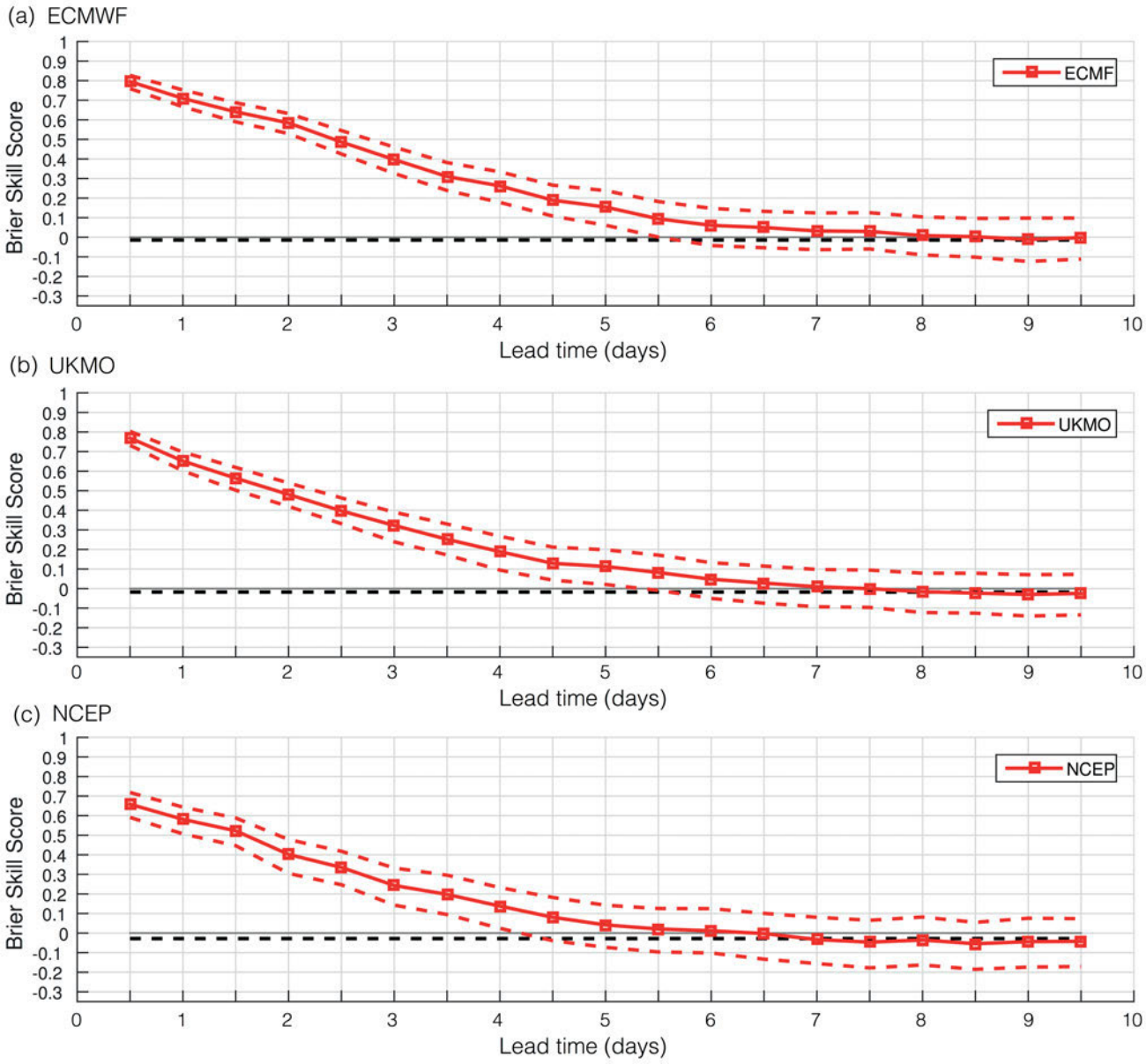


**Figure 8:** The Brier skill score (BSS) for extreme ramp forecasts constructed using the (a) ECMWF, (b) UKMO and (c) NCEP forecast systems. A ramp event is said to occur here if there is a GB wind power swing of at least 50 % of GB wind capacity within a 24 hour time window. Dashed lines indicate the 95 % confidence intervals, calculated using a bootstrapping technique. The BSS is calculated relative to a climatological prediction based on the analysis and 6 hour ahead control forecasts (as described in the text). For comparison, the relative skill of an equivalent climatological prediction based on ERA-Interim data from 1980–2013 is also shown (black, dashed).

The dependence of BSS on lead time for ramp forecasts produced by the ECMWF forecast system is shown in Figure 8(a). As for the upper bound calculated in Section 3.1, we use a threshold BSS = 0.1 to define the upper limit of predictability for probability forecasts ($t_P$), in order to avoid the issue of the skill remaining marginally positive as it asymptotes to zero. The skill drops below BSS = 0.1 on average after around 5.5 days, a considerably shorter limit of predictability than was found for the GB capacity factor forecasts (8 days). The 95 % confidence intervals for these statistics are relatively broad, reflecting the small number of extreme ramps contributing to the statistics. No robust seasonal variations in the BSS for ramps was found (not shown), though this may be due to the small sample size. For completeness, Figs. 8(b,c) show equivalent plots for the UKMO and NCEP forecast systems. Here, the limit of predictability for ramp forecasts is even shorter, at around 5 days ahead and 4.5 days ahead respectively.

## 4 Conclusions

This paper investigates the range of forecast lead times for which system-aggregated wind power predictions based on global numerical weather prediction (NWP) forecasts provide skilful information, using the Great Britain (GB) power system as an example. It is shown

**Table 2:** A summary of the values obtained for upper and lower bounds to the useful forecast range identified in this paper for different global ensemble forecast systems for the specific predictand of GB capacity factor ($t_U$ and $t_L$ respectively). The bottom row gives the limit to predictability associated with forecasting the probability of extreme power ramp events ($t_P$), defined by a change in GB capacity factor by more than 50 % within a 24 hour window centred on the lead time.

| | Approximate Lead Time (Days) | | |
|---|---|---|---|
| Bound | ECMWF | UKMO | NCEP |
| Upper ($t_U$) | 8 ± 1 | 7 ± 1 | 6 ± 1 |
| Lower ($t_L$) | 2.4 ± 0.2 | 1.9 ± 0.2 | 1.4 ± 0.2 |
| Ramp ($t_P$) | 5.5 (4.5–8) | 5 (4–7) | 4.5 (3.5–6.5) |

that they provide useful information within a window bounded at long lead times by a limit to predictability, beyond which the forecasts are no more skilful than a climatology-based statistical prediction. At short lead times, the total uncertainty is dominated by what we term the *unresolved uncertainty*, which is that associated with estimating the system-wide wind power response to the large-scale atmospheric state. The lead time beyond which the *resolved uncertainty* associated with atmospheric chaos (as represented by the global forecasts) is larger than the unresolved uncertainty determines the lower bound.

The upper and lower bounds on this range of lead times are summarised in Table 2. At long lead times, the upper limit of predictability for wind power forecasts at a specific point in time is found to vary seasonally. These forecasts retain skill relative to a climatology-based prediction out to around 7 days in summer but up to around 9 days in winter in the ECMWF forecast system. This upper bound was slightly lower for the UKMO and NCEP forecast systems.

The lower threshold, below which the *unresolved* uncertainty is found to be, on average, larger than the *resolved* uncertainty is around 1.4–2.4 days, depending on the forecast system. As the forecast lead time increases, improvements such as dynamical downscaling and microscale modelling (which act to reduce the *unresolved* uncertainty) will therefore be of diminishing value. State-of-the-art operational wind power forecast systems already employ more advanced downscaling and microscale modelling than is assumed here, and so the magnitude of the *unresolved* uncertainty reported in this paper is likely to be higher here than in operational forecast systems. This implies that, in practice, the resolved uncertainty could dominate at lead times below even 1.4–2.4 days.

The extreme ramp events considered here occur only rarely, and so the limit of predictability associated with extreme ramps are subject to much larger uncertainty, as measured using 95 % confidence intervals from bootstrap sampling. In this case, there are too few events to attribute this uncertainty to seasonal variability. These results suggest however that the limit of predictability associated with probability forecasts of extreme ramp events is significantly shorter than that associated with forecasts of wind power generation at a specific future point in time.

These results add to the understanding of system-aggregated wind power forecasts and the capabilities and limitations imposed on them by the underlying global NWP forecast systems. Wind power forecast developers and users should carefully consider the relative importance of the resolved and unresolved uncertainties when planning and resourcing new forecast developments, and be aware of the varying limits of predictability associated with different forecast metrics.

Furthermore, forecasting the probability of a wind power event is shown to be potentially more challenging than using an ensemble to forecast the system-wide wind power generation at a future time. For the extreme ramp example used here, the limit to predictive skill in the forecast of probability, $t_P$, is shorter than the $t_U$ estimated for GB wind power, but nevertheless considerably greater than the lower bound, $t_L$, showing that there is a wide window for which forecasts of probabilities can be useful for risk-based decision making.

## Acknowledgments

The authors would like to thank Dr Phil Coker (University of Reading, UK) for helpful discussions and for managing the relationship between the University of Reading and National Grid. This work was funded by National Grid who provided Great Britain wind power data and funding via Network Innovation Allowance grant NIA NGET0085. The authors would particularly like to thank David Lenaghan (National Grid) for helpful advice throughout this project and Andrew Richards (National Grid) for providing the observed wind power data. The authors also thank the two anonymous reviewers who provided helpful comments during peer review. Numerical weather prediction (NWP) data was retrieved using the Meteorological Archival and Retrieval System (MARS) at the European Centre for Medium Range Weather Forecasts (ECMWF). The authors are particularly grateful to Paul Dando (ECMWF) for technical help with the NWP data retrieval. All plots were produced using MATLAB, and Figure 1 made use of the M_MAP mapping package v1.4f, courtesy of Prof. Rich Pawlowicz (University of British Columbia, Canada; http://www.eos.ubc.ca/~rich/map.html).

## References

Bougeault, P., Z. Toth, C. Bishop, B. Brown, D. Burridge, D.H. Chen, Co-authors, 2010: The THORPEX interactive grand global ensemble. – Bull. Amer. Meteor. Soc. **91**, 1059–1072.

Brier, G., 1950: Verification of forecasts expressed in terms of probability. – Mon. Wea. Rev. **78**, 1–3.

Cannon, D.J., D.J. Brayshaw, J. Methven, P.J. Coker, D. Lenaghan, 2015: Using reanalysis data to quantify extreme wind power generation statistics: a 33 year case study in Great Britain. – Renew. Energ. **75**, 767–778.

Carrillo, C., A.F. Obando Montaño, J. Cidrás, E. Díaz-Dorado, 2013: Review of power curve modelling for wind turbines. – Renew. Sustain. Energy Rev. **21**, 572–581.

Cutler, N., M. Kay, H. Outhred, I. MacGill, 2007: High-risk scenarios for wind power forecasting in Australia. Proceedings of the European Wind Energy Conference and Exhibition (EWEC), Milan, Italy. – Available online at http://www.ewea.org/ewec2007/allfiles2/102_Ewec2007fullpaper.pdf.

Dee, D.P., S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, others, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. – Quart. J. Roy. Meteor. Soc. **137**, 553–597.

Drew, D.R., D.J. Cannon, D.J. Brayshaw, J.F. Barlow, P.J. Coker, 2015: The impact of future offshore wind farms on wind power generation in Great Britain. – Resources **4**, 155–171.

Ferreira, C., J. Gama, L. Matias, A. Botterud, J. Wang, 2010: A survey on wind power ramp forecasting. – Tech. Rep. ANL/DIS-10-13, Argonne National Laboratory.

Frame, T.H.A., M.H.P. Ambaum, S.L. Gray, J. Methven, 2011: Ensemble prediction of transitions of the north atlantic eddy-driven jet. – Quart. J. Roy. Meteor. Soc. **137**, 1288–1297.

Giebel, G., R. Brownsword, G. Kariniotakis, M. Denhard, C. Draxl, 2011: The state-of-the-art in short-term prediction of wind power: A literature overview. – Available online at http://orbit.dtu.dk/fedora/objects/orbit:83397/datastreams/file_5277161/content.

Glahn, H.R., D.A. Lowry, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. – J. Appl. Meteor. **11**, 1203–1211.

Gray, S.L., C.M. Dunning, J. Methven, G. Masato, J.M. Chagnon, 2014: Systematic model forecast error in Rossby wave structure. – Geophys. Res. Lett. **41**, 2979–2987.

Greaves, B., J. Collins, J. Parkes, A. Tindal, 2009: Temporal forecast uncertainty for ramp events. Proceedings of the European Wind Energy Conference and Exhibition (EWEC), Marseille, France. – Available online at http://proceedings.ewea.org/ewec2009/zipfilespdf/Technical.zip under file name "CT4_2.pdf".

GWEC, 2015: Global wind report: Annual market update 2014. Global Wind Energy Council (GWEC). – Available online at http://www.gwec.net/publications/global-wind-report-2.

Hagedorn, R., R. Buizza, T.M. Hamill, M. Leutbecher, T.N. Palmer, 2012: Comparing TIGGE multi-model forecasts with reforecast-calibrated ECMWF ensemble forecasts. – Quart. J. Roy. Meteor. Soc. **138**, 1814–1827.

Hersbach, H., 2000: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. – Wea. Forecast. **15**, 559–570.

Jiménez, P.A., J.F. González-Rouco, E. García-Bustamante, J. Navarro, J.P. Montávez, J.V.-G. de Arellano, J. Dudhia, Antonio Muñoz-Roldan, 2010: Surface wind regionalization over complex terrain: Evaluation and analysis of a high-resolution WRF simulation. – J. Appl. Meteor. Climatol. **49**, 268–287.

Johnson, C., R. Swinbank, 2009: Medium-range multi-model ensemble combination and calibration. – Quart. J. Roy. Meteor. Soc. **135**, 777–794.

Lorenz, E.N., 1963: Deterministic nonperiodic flow. – J. Atmos. Sci. **20**, 130–141.

Lorenz, E.N., 1975: Climatic predictability. – In: The physical basis of climate and climate modelling, 132–136, World Meteorological Organisation (WMO) GARP publication series, Vol. 16. – Available online at http://library.wmo.int/pmb_ged/garp-ps_16.pdf.

Lynch, K.J., D.J. Brayshaw, A. Charlton-Perez, 2014: Verification of European subseasonal wind speed forecasts. – Mon. Wea. Rev. **142**, 2978–2990.

Mass, C.F., D. Ovens, K. Westrick, B.A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? – Bull. Amer. Meteor. Soc. **83**, 407–430.

Morales, J.M., A.J. Conejo, H. Madsen, P. Pinson, M. Zugno, 2014: Integrating renewables in electricity markets – operational problems, volume 205. – Springer.

Pinson, P., R. Hagedorn, 2012: Verification of the ECMWF ensemble forecasts of wind speed against analyses and observations. – Meteor. Appl. **19**, 484–500.

Reid, S.J., R. Turner, 2001: Correlation of real and model wind speeds in different terrains. – Wea. Forecast. **16**, 620–627.

Richardson, D.S., 2012: Chapter 9: Economic value and skill. – In: Jolliffe, I.T., D.B. Stephenson (Eds.): Forecast verification: A practitioner's guide in atmospheric science, Second edition. – Wiley-Blackwell.

Rienecker, M.M., M.J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, others, 2011: MERRA – NASA's Modern-Era Retrospective analysis for Research and Applications. – J. Climate **24**, 3624–3648.

Swinbank, R., M. Kyouda, P. Buchanan, L. Froude, T. Hamill, T. Hewson, J. Keller, M. Matsueda, J. Methven, F. Pappenberger, M. Scheuerer, H. Titley, L. Wilson, M. Yamaguchi, 2016: The TIGGE project and its achievements. – Bull. Amer. Meteor. Soc. **97**, 49–67, DOI:10.1175/BAMS-D-13-00191.1.