Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 76 (2015) 216 – 222

Energy Procedia

European Geosciences Union General Assembly 2015, EGU

Division Energy, Resources & Environment, ERE

# Operational Evaluation of a Wind-Farm Forecasting System

Enda O'Brien*, Adam Ralph

*Irish Centre for High-End Computing, IT Building, NUI Galway, Ireland*

**Abstract**

Performance of a wind-forecasting system for a wind-farm in Ireland is reported. Forecasts were based on ensembles constructed from HARMONIE model runs every 6 hours, along with extra high-resolution HARMONIE runs every 12 hours. Statistical post-processing with Bayes Model Averaging (BMA) removed bias very effectively. The "raw" incremental skill provided by each extra ensemble member was negligible, but the net value, after BMA post-processing, was significantly larger. Thus, a small ensemble with BMA is more skillful than a larger ensemble with simple averaging only. A larger ensemble is still more skillful than a smaller one, if both use BMA.


Peer-review under responsibility of the GFZ German Research Centre for Geosciences

*Keywords:* Bayes Model Averaging; Wind-farm; Mean Absolute Error

## 1. Introduction

This article reports on the ability and skill of a prototype forecasting system, developed at the Irish Centre for High-End Computing (ICHEC), to make routine, fully automated wind and power forecasts for each of the 4 x 2.3MW turbines in a wind-farm on mountainous terrain in southwest Ireland.

* Corresponding author. Tel.: 353-91-495946.
*E-mail address:* enda.obrien@ichec.ie


Peer-review under responsibility of the GFZ German Research Centre for Geosciences
doi:10.1016/j.egypro.2015.07.903

The main component of the system is the HARMONIE Numerical Weather Prediction (NWP) model [1]. An "ensemble" of several different model runs (using different configurations, or starting at different times) all contribute to the final forecast.

The second key component is the "ensemble-BMA" (or Bayes Model Averaging) package from the R statistical programming language [2,3]. This package measures how the interpolated forecasts verify against observations over a past "training set" (e.g., the prior 20 days) and tries to detect any systematic errors, or biases, from such comparisons. These biases can be used to assign different weights to each member of an ensemble, and are ultimately removed from a "final" forecast. Although each final or "best guess" forecast is fundamentally obtained from physical principles as expressed in the NWP model, it is also adjusted to take into account what can be learnt from the statistics of previous performance.

The HARMONIE domain used for most forecasts covers a region slightly larger than Ireland and the UK (approx. 1,200 x 1,500 $km^2$), has a horizontal grid-resolution of 2.5 km, 65 vertical levels, and is run (operationally, by Met Éireann) 4 times daily at 00Z, 06Z, 12Z and 18Z. The first 30 hours of each forecast run were used for the wind-farm forecasting system. HARMONIE is "nested" in a larger but coarser-resolution global model. In other words, its boundary conditions are taken from the global model.

Output from operational HARMONIE runs were kindly provided by Met Éireann for use by this project. As long as such permission is granted, accessing the daily output files is straightforward, since the Met Éireann forecasts are run on the ICHEC "Fionn" supercomputer, and output remains on Fionn for several days before being rotated out to archive at Met Éireann.

Specifically for the purposes of this project, HARMONIE was also configured with a finer resolution (0.5 km) over a smaller domain centred on Kerry, and "nested" in turn inside the 2.5 km HARMONIE. In other words, the 0.5 km model takes its boundary condition updates from the 2.5 km model. This model was run twice daily (starting at 00Z and 12Z) to produce forecasts out to 30 hours, with shorter 6-hour forecasts starting at 06Z and 12Z to ensure the smooth "blending" of initial fields from one run to the next.

The wind-farm forecasting system reported on here is very similar to that described in [4]. In the present case it is applied to a different wind-farm, and is used to make real-time (operational) forecasts in a fully automated way. The system was also adapted to make historical "hind-casts" as well, using several different ensemble constructions from NWP model output. The main new contribution of this paper is to show some real-time "products", along with a simple evaluation of system performance. This shows the relatively large contribution to forecast skill made by BMA, and the relatively small contribution made by each incremental ensemble member.

## 2. Data

Observed wind-speeds at each individual turbine, along with the power generated, were provided from 1st Jan. 2014 until 8th March 2015. Data up to 31st Dec. 2014 was "historical" and only useful for verification of "hind-casts". Observational data provided after 1st Jan. 2015 lagged the forecasts, which were made in "real-time" (though of course those data can now also be used for "hind-casts").

While the turbine operator provided observational data every 10 minutes, only values at the start of each hour were used for forecast verification purposes, in order to correspond with forecast output which was only available at hourly intervals. Thus, 5/6 of the information contained in the observations was not used at all. This represents quite a large waste of data, and is an issue worth re-visiting in the future.

Courtesy of Met Éireann, archived forecasts from the 00Z and 12Z runs of the operational HARMONIE model were provided from 1st Jan. to 31st Oct. 2014, at "standard" pressure levels (mean sea-level, 850 hPa, 700 hPa, and higher). The sequence of forecasts made by those twice-daily runs formed the basis of a ***2-member "ensemble"***, for the purposes of BMA processing.

From 1st Nov. 2014 onwards, more complete output was saved from all 4 operational HARMONIE forecast runs each day, and from all 65 model levels. Those runs formed the basis of a ***4-member "ensemble"*** for BMA purposes.

Also, from 1st Nov. 2014 onwards, complete output was also available from the 00Z and 12Z runs of the high-resolution 0.5km HARMONIE forecasts nested inside the 2.5km operational runs. When combined with the 4 operational runs each day, these constituted a ***6-member "ensemble"*** for BMA purposes.

## 3. Ensemble Construction

While BMA expects to perform its averaging over an ensemble of forecasts, no true ensemble was run, at least not in the standard sense of several simultaneous forecast runs differing only slightly in initial conditions or physical configuration. Instead, for BMA training purposes, all the forecasts started at, say, 00Z each day were concatenated into a continuous forecast stream and designated as a single member of the "ensemble", with the joining "seams" occurring every 24 hours, even though each individual run was for at least 30 hours. Other analogous ensemble members were generated by concatenating forecasts starting at 06Z, 12Z or 18Z, respectively. In this way, output from an overlapping sequence of forecast runs can be used to generate a "poor-man's ensemble" in a way that is amenable to processing by the BMA package.

When it comes to making a real live forecast, such sequences of ensemble members always have "ragged" or "staggered" endings, with one member extending to 30 hours, with the others finishing 24, 18, or 12 hours into the future. The ensemble BMA package, however, has no problem with this, and simply treats the hours past the end of each truncated sequence as "missing data" in a sensible way. While other ensemble constructions are possible, this staggered ending to each forecast sequence is simply in the nature of operational BMA forecasting: there will always be one run that is the "most recent", and which will necessarily have *all* the weighting during the latter part of the forecast period.

Ideally, all 30 hours of each forecast run should be saved and used in the "training set". Ideally too, different weights should be given to the "most recent" forecast, and the others, which are "at least 6-hours", "at least 12-hours" and "at least 18-hours" old. There is a lot of flexibility in assigning ensemble "weights" in the R program's "ensembleBMA" package, and only one option has been tested so far. Nevertheless, it is difficult to see how all those conditions can be met in the context of "ensembleBMA", since it seems to require each ensemble member to be a single continuous time-series. So in practise, the last 6 hours of each 30-hr forecast are over-written by output from a new run started 24 hours later.

At 09Z each day (when the BMA analysis is currently run), output from only one ensemble member (the 06Z run of the 2.5km model) is available out to 12Z the following day. Meanwhile, output from 3 members (the 06Z run of the 2.5km, and the 00Z runs of the 2.5km and 0.5km models), are available out to 06Z the following day. Forecast contributions from all 6 ensemble members are available out to midnight that night. For the first 18 hours of each BMA "averaging" then, more weight really ought to be given to those members started most recently. Currently all members are treated equally in that regard; BMA does not "know" which ensemble member is "most recent" – only that some members have " missing values" later into the forecast.

One further aspect, which remains somewhat arbitrary, is the optimal length of the BMA "training period. So far we have not had the opportunity to test this systematically. However, tests using both 100 days (the longest possible with the data currently available) and 20 days suggest that the results are not really sensitive to this training period length – although run-times are significantly longer for the 100-day period!

## 4. Results

An example of the "product" provided to the wind-farm operator on a daily basis is shown in Fig. 1. Charts like these were automatically generated and sent at approx. 9:00 am each day, as soon as the 06Z forecast run and subsequent BMA post-processing was completed. The left panel shows wind-speed, the right panel power output, for one particular turbine. Power forecasts were derived entirely from wind-speed forecasts using a standard power curve for each turbine. The real work is in forecasting the wind; the power forecast is then obtained at the final step in the process as a simple (but non-linear) function of wind-speed.

In each panel of Fig. 1, the black curves show the verifying observed value. These values of course are not available when the forecast is made – they were added when available later on for validation purposes. The blue curve in each panel of Fig. 1 represents the raw (6-member) ensemble-mean forecast; the solid red curve is the BMA "expected" forecast (i.e., the 50% quantile), while the lower and upper dashed red curves are the 20% and 80% quantiles, respectively, forecast by BMA (all based on the 6-member ensemble). In other words, the actual verifying wind-speed (or power output) is expected to be below the 20% quantile approx. 20% of the time, and above the 80% quantile approx. 20% of the time. Verifying observations are expected to fall within the "envelope" bounded by the dashed red curves about 60% of the time. If the separation between the dashed red curves is small, it reflects the fact that ensemble member winds all cluster closely around the mean, and these forecasts can be made with higher confidence than when the distance between the dashed red curves is large.

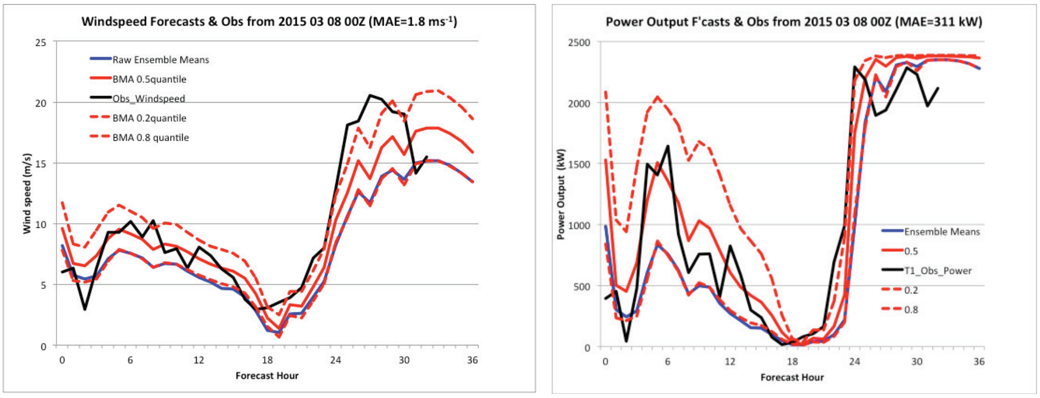


Fig. 1.Wind-speed (left panel) and power output (right panel) from a day with a "ramp" event. See text for details.

It is clear from Fig. 1 that the raw ensemble members (blue curve) have a negative bias in the forecast wind-speed of about 1 $ms^{-1}$ for this turbine, which BMA post-processing is able to remove (based on the prior 20-day training period).

A noteworthy feature of Fig. 1 is that it shows a "ramp" event, or sudden increase in wind-speed (and associated power output). The forecast captures this reasonably well, though it underestimates somewhat the onset speed and final magnitude (over 20 $ms^{-1}$) of the event. Note that no verifying observations were provided after 32 hours (i.e., the turbine showed a non-zero error "status" after this time) – possibly because it shut down in the high winds. The wind-speed mean absolute error (MAE) is 1.8 $ms^{-1}$, while the power output MAE is 311 KW, over those 32 hours.

Forecast charts (without the verifying observations) as in Fig. 1 were provided for each turbine, as well as for the entire farm as a whole. The systematic bias in the raw ensemble forecasts was negative for some turbines, while it was positive for others (located up to 1 km away, on a hill slope with a completely different orientation). Not surprisingly, the MAE for the farm-wide forecasts was typically lower than for each individual turbine, since the individual errors from the "pin-point" individual forecasts tended to cancel each other out.

For evaluation purposes, individual forecasts for each turbine (as in Fig. 1) were collected into sets of 100 days in length, and BMA analyses were run on these in "hind-cast" mode. The MAE was calculated from the ensemble members at each hour, and then averaged for each day, and plotted as shown in Figs. 2-3. Fig. 2 shows such a 100-day sequence from the 2-member ensemble from 8$^{th}$ March 2014; the daily mean wind-speed for each day for one particular turbine is shown as the black, along with the daily-mean MAEs from the raw ensemble-mean forecasts (red curve) and daily-mean MAEs from the BMA forecasts (green curve).

Fig. 3 shows 100 days from 29$^{th}$ Nov. 2014, with observed wind-speeds again as the black curve; the "raw" ensemble MAEs from the 4- and 6-member ensembles shown as the blue curves (virtually indistinguishable); BMA-generated MAEs from the 4-member ensemble as the red curve, with BMA-generated MAEs from the 6-member ensemble as the green curve.

It is clear from both Figs. 2 and 3 that the BMA post-processing is able to reduce the MAE from the raw ensemble mean almost all the time. In that regard, BMA provides significant extra forecasting skill. It is also apparent in Fig. 3, however, that if only the "raw" ensemble means were used for forecasting purposes, then the extra 2 ensemble members do not appear to provide any extra skill at all. However, BMA is able to use the extra 2 ensemble members to reduce the final BMA forecast errors very slightly. In all cases, the forecast errors tend to be larger when the wind-speed itself is stronger.

These results are summarized in Table 1, which attempts to distil the forecasting skill of each model ensemble down to a single wind-speed MAE. The variability of the MAEs among the 3 different 100-day periods with the 2-member ensemble provides some sense of the "natural" variability within the system, although the reduction in errors during the summer months also reflects the general drop in background wind-speed during that period. As expected, in all cases the BMA MAEs are smaller than the raw ensemble-mean MAEs.

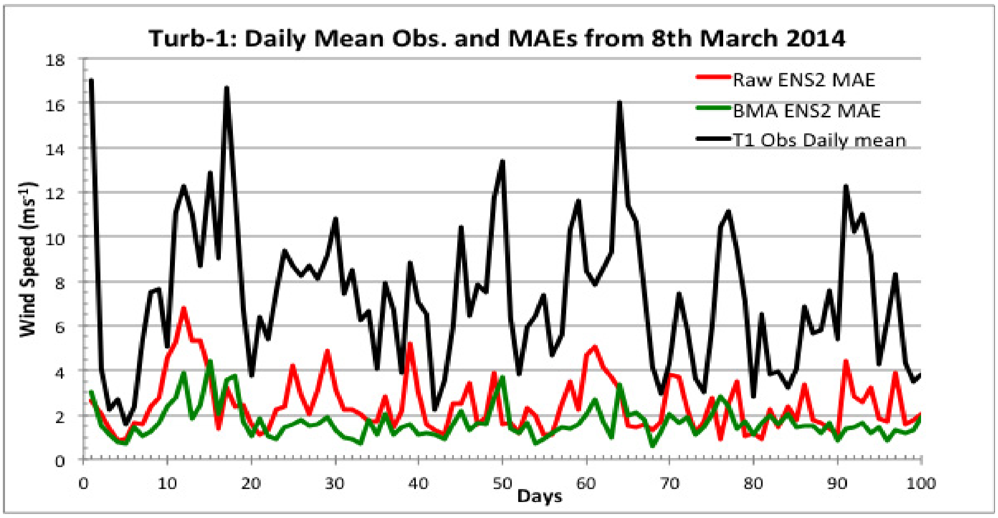


Figure 2.Full observed wind-speed (black curve) along with MAEs from 2-member ensemble averages (raw & BMA).

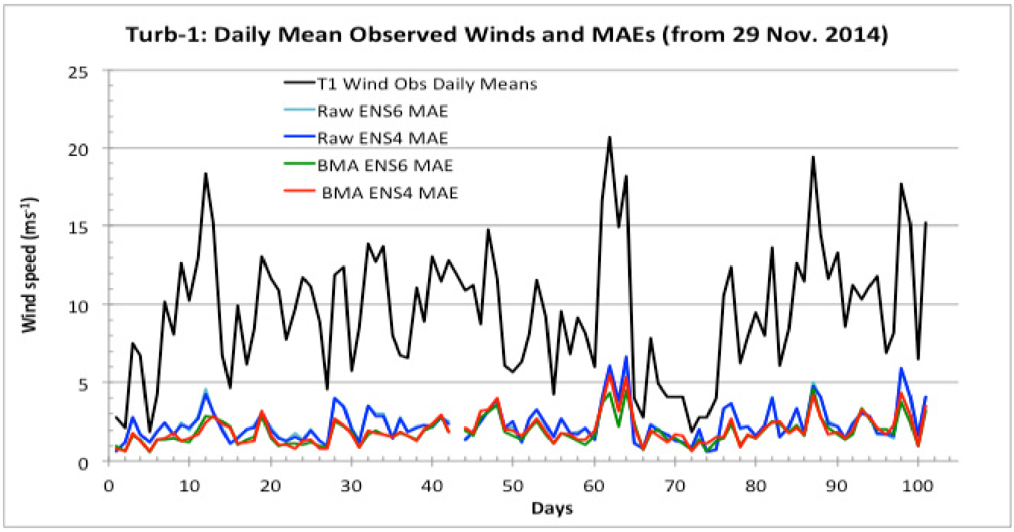


Figure 3.Full observed wind-speed (black curve) along with MAEs from 4- and 6-member ensembles.

Also shown in Table 1 are the MAEs as a percentage of the mean background wind-speed during each of the 100-day periods. Thus, even though the smallest absolute MAE (1.48 $ms^{-1}$) was obtained with the minimal and crude 2-member ensemble, this occurred during the relatively calm summer months (from 16 June, 2014). In relative terms (i.e., MAE as a fraction of the background wind-speed), the best result was obtained from the most complete 6-member ensemble, where the MAE of 1.88 $ms^{-1}$ represented just 22% of the observed background wind-speeds.

The crudeness of the 2-member ensemble may be seen in the very high fractional errors of the raw ensemble-mean forecasts, which are over 35% of the full wind-speed in all cases. This seems to be related to the difficulty in interpolating to turbine hub height from the winds at the surface and at 850 hPa, which is usually the next available standard pressure level above the surface. For the 4- and 6-member ensembles, in contrast, output was available at all model levels, at least 4 of which were within the boundary layer, making interpolation of resolved winds to turbine height significantly more accurate.

Table 1. Summary of MAEs in wind-speed forecasts for one particular turbine for 100-day periods.

| Ensemble Members | Time Period (each 100 days) | MAE ($ms^{-1}$) (raw winds) | MAE ($ms^{-1}$) (BMA winds) |
|---|---|---|---|
| 2 | From 8 Mar. 2014 | 2.46 (37% of obs. wind) | 1.66 (26% of obs. wind) |
| 2 | From 16 Jun. 2014 | 2.23 (41% of obs. winds) | 1.48 (28% of obs. wind) |
| 2 | From 24 Sep. 2014 | 2.59 (35% of obs. wind) | 1.80 (24% of obs. wind) |
| 4 | From 29 Nov. 2014 | 2.34 (25% of obs. wind) | 1.95 (23% of obs. wind) |
| 6 | From 29 Nov. 2014 | 2.35 (25% of obs. wind) | 1.88 (22% of obs. wind) |

It should be noted that all errors are calculated as the difference between forecast and observed values at a particular turbine and a precise time. While the curves shown in Figs. 1-3 and the numbers in Table 1are all for just one turbine, they are quite representative of the other three turbines on the wind-farm, and also representative of the farm-wide (averaged) wind-speeds. Such "pin-point" forecasts inevitably incur relatively large errors since topography-generated waves or other gravity waves that are not represented exactly in the models can lead to large fluctuations in wind-speed, and thus large errors, as shown in the figures above. Errors computed as the difference between modelled and observed winds averaged over a 10 or 20-minute window, or averaged over the horizontal extent of the wind-farm, would be expected to be smaller. Indeed, when the MAEs are calculated on a farm-wide basis instead of for just one turbine, the MAEs on the last row of Table 1 decrease from 2.35 to 1.60 $ms^{-1}$ (for the raw ensemble means) and from 1.88 to 1.54 $ms^{-1}$ (for the BMA ensemble means), respectively.

Fig. 4 is analogous to Fig. 3, but with MAEs calculated on a farm-wide basis, i.e., where the verifying observations were the mean wind-speeds over all four turbines. Fig. 4 also shows MAEs from just the 6-member ensemble (i.e., without the 4-member ensemble as well, as in Fig. 3). Otherwise, the time-span and the scale on the vertical axis are the same in both cases. As might be expected, the averaging process reduces the extremes in both observed wind-speed and also in the MAEs that appear in Fig. 3. (E.g., compare days 62-64, or day 99, in both Figs. 3 and 4). Moreover, as mentioned above, the overall errors are slightly smaller when computed from the spatial average of all the turbines instead of from the "pin-point" wind-speeds at a single turbine.

## 5. Discussion

When making wind-speed forecasts for wind-farms, simple interpolation from the output of a minimal 2-member ensemble constructed from the operational runs of a standard NWP model generates very large errors (35-40% of the true wind-speed). However, a statistical post-processing package like BMA has the ability to greatly reduce those errors to approx. 25-30% of the true wind-speed. Operating a wind-farm forecasting system based on just those components is relatively cheap (both computationally and financially). It is possible to achieve even better accuracy by the addition of more ensemble members. The evidence from our tests with 2-, 4- and 6-member ensembles (as summarized in Table 1) is that and the addition of each incremental pair of ensemble members does indeed reduce the MAEs, at least after the raw ensemble output has been post-processing with BMA. However, such error reductions "at the margins" are relatively small and incremental. Depending on the value attributed to accurate forecasts, the incremental skill provided by the last couple of ensemble members may not justify the extra costs involved in running and post-processing them. Of course, the ever increasing costs involved in obtaining ever more incremental improvements in forecast skill are characteristic of NWP in general.

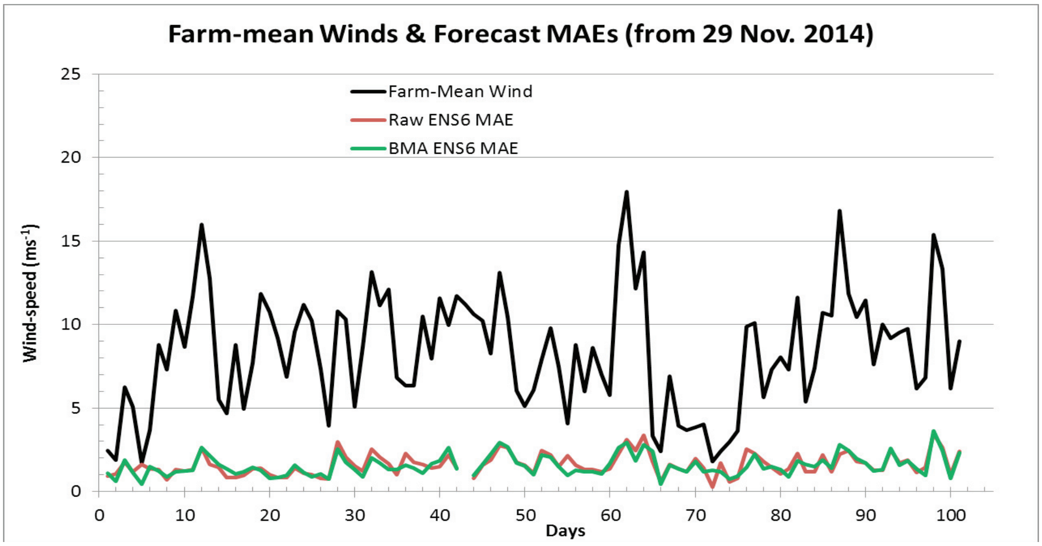


Figure 4 .Observed winds and MAEs as in Fig. 3, except based on a farm-wide average, and not showing ENS4 MAEs.

## Acknowledgements

We are grateful to Enterprise Ireland for funding this work. We also thank Met Éireann and Mainstream Renewable Power for use of their data.

## References

[1] Seity, Y, Brousseau P, Malardel S, Hello G, Bénard P, Bouttier F, Lac C, Masson V. The AROME-France Convective-Scale Operational Model. *Mon. Wea. Rev*. 2011, **139**, 976-991.

[2] Raftery AE, Gneiting T, Balabdaoui F, Polakowski M. Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Mon. Wea. Rev*. 2005, **133**, 1155–1174.

[3] Gneiting T, Sloughter JM, Raftery AE. Probabilistic wind speed forecasting using ensembles and bayesian model averaging. *J. Amer. Statistical Assn*., 2010, **105**, 25-35.

[4] Peters, M.B., E. O'Brien, A. McKinstry, and A. Ralph, 2013. Wind Forecasting using HARMONIE with Bayes Model Averaging for Fine-tuning. Energy Procedia, **40**: 95-101. http:/www.sciencedirect.com/science/article/pii/S1876610213016068.