Machine Learning Based Multi-Physical-Model Blending for Enhancing Renewable Energy Forecast – Improvement via Situation Dependent Error Correction.*

Siyuan Lu[†, ξ], Youngdeok Hwang[†], Ildar Khabibrakhmanov[†], Fernando J. Marianno[†], Xiaoyan Shao[†], Jie Zhang[‡], Bri-Mathias Hodge[‡], Hendrik F. Hamann[†]

[†]*IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA*

[‡]*National Renewable Energy Laboratory, Golden, CO 80401, USA*

***Abstract*—With increasing penetration of solar and wind energy to the total energy supply mix, the pressing need for accurate energy forecasting has become well-recognized. Here we report the development of a machine-learning based model blending approach for statistically combining multiple meteorological models for improving the accuracy of solar/wind power forecast. Importantly, we demonstrate that in addition to parameters to be predicted (such as solar irradiance and power), including additional atmospheric state parameters which collectively define weather situations as machine learning input provides further enhanced accuracy for the blended result. Functional analysis of variance shows that the error of individual model has substantial dependence on the weather situation. The machine-learning approach effectively reduces such situation dependent error thus produces more accurate results compared to conventional multi-model ensemble approaches based on simplistic equally or unequally weighted model averaging. Validation over an extended period of time results show over 30% improvement in solar irradiance/power forecast accuracy compared to forecasts based on the best individual model.**

I. INTRODUCTION

Accurate forecasts of the atmospheric state have been a particularly challenging problem yet have enormous social and economic benefits [1]. Among many important applications of atmospheric forecasts, the need for the forecast of variable solar and wind energy generation is becoming pressing with increasing penetration of solar and wind energy to the total energy mix. In the arena of solar forecasting, significant progresses towards minute-, hour-, and day-ahead forecast have been made recently utilizing an array of methodologies based on sky-camera imagery, satellite imagery, numerical weather predictions, and/or statistical analysis [2-8]. Forecasts of enhanced accuracy are highly desired for reliable grid operation at reduced cost and a more efficient electricity market [9-13].

*Research is partially supported by US Department of Energy Contract #DE-EE0006017

[ξ] Corresponding Author Phone: 1-914-945-1809;
e-mail: lus@ us.ibm.com (Siyuan Lu).

Similar to other complex system problems, some of the major milestones towards better prediction skills for atmospheric states are achieved by cleverly combining physical and statistical modeling approaches. Examples include Kalman filtering for data assimilation [14,15], model output statistics approach for deriving site specific forecast [16,17], and ensemble averaging of multiple models [18]. Pertaining to the latter, it was demonstrated that averaging of ensemble members using unequal weight coefficients obtained by regressing historical data – the so-called super ensemble approach – often further enhances accuracy, compared to simplistic ensemble averaging [19,20].

The cornerstone for these methods of combining physical and statistical approaches is, by-and-large, the fundamental Bayes Theorem. In parallel, exploding progresses were made in the supporting infrastructure, including the advent of Big Data and advanced computing infrastructures [21,22] which eased data access and organization as well as the development of efficient and robust machine learning algorithms [23, 24]. In this article we show that compared to conventional multi-model approach, further enhanced accuracy can be achieved if we combine the individual model forecasts using a machine-learning based approach which explicitly takes into account appropriately chosen additional state parameters beyond these need to be explicitly forecasted (Figure 1). For instance, even if we need to predict only solar irradiance, the machine learning algorithm needs to take as input also parameters such as solar zenith angle, column integrated cloud water content etc. meaning that different models provide different skills based on these additional state parameters. As discussed below, these additional parameters collectively create different weather categories in which the different models exhibit different error characteristics. Thus the machine-learning algorithm, via effectively combining and correcting the individual models differently in the different categories, may be expected to achieve better overall accuracy. With such an approach, we demonstrate ~30% accuracy improvement, as measured in terms of root mean square error (RMSE) or absolute mean error (MAE), for the forecast of meteorological parameters (solar irradiance, wind speed, temperature) as well as PV power generation.

In the following we present (a) the rationale behind our multi-model blending approach – the weather category dependent model error characteristic – analyzed using functional analysis of a variance (FANOVA, Sec. II.A) [25]; (b) the implementation of the multi-model blending (Sec. II.B); (c) case studies illustrating the improved forecast accuracy for both meteorological parameters and PV power enabled by the model blending approach (Sec. III); (d) results supporting the postulation that the machine-learning based model blending improves forecast accuracy via situation dependent error correction (Sec. IV); and (e) conclusions (Sec. V).

II. Machine Learning Based Multi-Model Blending

A. Physical model error dependence on weather situation category.

Limited by the simplification and coarse-graining of the actual atmospheric system, the raw numerical weather prediction outputs often has significant systematic error. As an illustrative example, we analyze the error of one of the operational numerical weather prediction model NAM (North American Mesoscale Model, 5 km resolution) operated by national oceanic and atmospheric administration (NOAA) [26] and how the error depends on the weather situation category.

For this analysis, measurements of global horizontal solar irradiance (GHI) are taken from NOAA's seven first-class measurement stations of the surface radiation (Surfrad) network [27] for period from 2013-07-05 to 2014-01-21. The GHI forecast of hourly resolution as well as additional atmospheric parameters such as direct normal irradiance (DNI), column integrated cloud liquid water (atmosphere treated as a single layer) and cloud ice contents, cloud base and top heights, temperature (2 m above ground), ground pressure, wind speed (10 m above ground), surface rain etc. are derived from the daily run of the NAM model at 06z. Feeding these input data (24 hors x 201 days x 7 sites = 33768 sets of measurements of GHI as well as forecasts of GHI and other parameters) into a quantile forest regression model [28, 29], one can fit of the relationship between the GHI forecast error and the forecasted parameters,

$$E_{GHI}=F(x_1, x_2, \ldots, x_n), \tag{1}$$

where $E_{GHI} = GHI_{forecasted} - GHI_{measured}$, is the GHI forecast error, x_1 is $GHI_{forecasted}$, x_2, …, x_n are the additional forecasted input parameters.

The errors of the forecast are then broken to its 0th, 1st, 2nd … order dependence on individual input parameters using functional analysis of variance (FANOVA) [25] Equ.(2)-(5). In essence, FANOVA decomposes the overall error into mean bias, the dependence on individual variables, the interaction between two-variable pairs etc. The zeroth order term f_0 (Equ. (3)) is the mean bias error of a forecast. The first order term f_i (Equ. (4)) provides the error dependence on x_i only while the effects of all other parameters are averaged out (with zeroth order term removed). The second order term $f_{i,j}$ (Equ.(5)) provides the error dependence on x_i and x_j (with zeroth and first order terms removed):

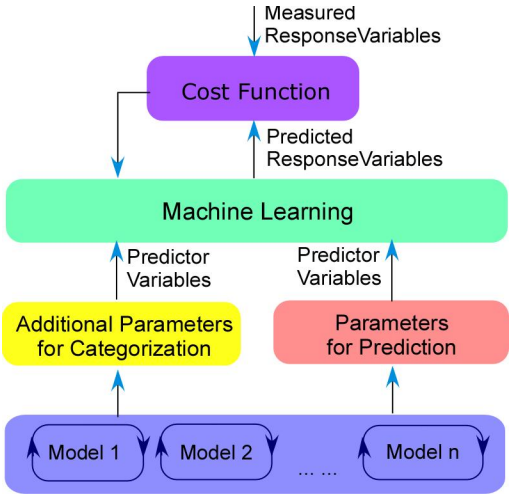

Figure 1.General schematic showing the architecture of the machine learning based model blending.

$$F = f_0 + \sum_i f_i(x_i) + \sum_{i\neq j} f_{i,j}(x_i, x_j) + \ldots \tag{2}$$

where,

$$f_0 = \int F(x_1,\ldots,x_n)\,dx_1 \ldots dx_n \tag{3}$$

$$f_i = \int F(x_1,\ldots,x_n)\,dx_1 \ldots dx_{i-1}dx_{i+1}dx_n - f_0 \tag{4}$$

$$f_{i,j} = \int F(x_1,\ldots,x_n)\,dx_1 \ldots dx_{i-1}dx_{i+1}\ldots dx_{j-1}dx_{j+1}\ldots dx_n - f_i(x_i) - f_j(x_j) - f_0 \tag{5}$$

Thus a FANOVA term having large dependence on the input variables means that the systematic error of the forecast is strongly dependent on situation category, vice versa. Plotted in Fig. 2A is a summary of the estimated first order FANOVA terms (f_i) of NAM GHI (first order error dependence on all input parameters) calculated using data from all seven Surfrad sites. Note that for plotting in Fig 2A, all input parameters have been linearly rescaled to [0,1] from their corresponding minimum and maximum values so that they can be overlaid on the same plot, and the results are smoothed to show the pattern more clearly. The result shows that the NAM GHI forecast error has a strong dependence on GHI, zenith angle, and cloud liquid water inputs. A negative 1st order error (under-prediction) is evident for small GHI or small zenith angle while a positive 1st order error (over-prediction) occurs for large GHI or zenith angle. Moreover, a positive 1st order error occurs for near-zero cloud liquid water content. Beyond first order, examples of 2nd order GHI forecast error dependence on input parameters are shown in Fig. 2B and Fig. 2C. We observe that GHI forecast error vs. GHI and zenith angle (Fig. 2B) can be divided into four regions. For small (large) zenith angle and small (large) GHI, the 2nd order forecast error is negative, otherwise the 2nd order error is

positive. As a second example, there is also a stronge interaction between column integrated cloud liquid water content and zenith angle (Fig. 2C). Large positive and negative 2nd order error occurs when there is near zero cloud liquid water content. We suggest that such 1st and 2nd order FANOVA GHI forecast error dependences represent a systematic error in the NAM modeling, potentially related to the relatively simplistic cloud microphysics and radiative transfer schemes employed.

B. Machine-learning based multi-model blending.

Similar to the NAM model, two other state-of-art numerical weather prediction models, rapid refresh (RAP, formerly Rapid Update Cycle) [30] and high-resolution rapid refresh (HRRR) [31] are also analyzed using FANOVA. It is apparent that these models have different manifestations of error dependence on the input parameters. Such observation suggests that if one divides the total space formed by those parameters on which the error depends on strongly into subspaces, as illustratively marked by dashed lines on Fig 2B, 2C, one may be able to reduce the errors of the individual models more effectively.

How to divide the space of the input parameter is often too complicated to be tackled by manual analysis, which prompted us to deploy a machine learning based approach as schematically depicted in Fig. 1. From global weather forecast models (such as the GFS – Global Forecast System) and observational inputs such as sounding measurement of vertical profile of atmospheric pressure, temperature, humidity etc., multiple physical models (blue box) such as NAM, RAP, HRRR etc. employ different configurations of grid setting and microphysics packages to forecast atmospheric states. Among the output of the models, some are of direct interest for prediction, such as solar irradiance, wind speed, temperature etc. for solar and wind power forecast. Those parameters are referred to below as prediction parameters (red box). Additionally the physical models also predict parameters such as cloud liquid water and ice contents etc., which are referred to below as categorization parameters (yellow box in Fig. 1). These parameters are of no direct interest for the intended application and there is no corresponding measurement to validate the forecasts of these parameters. The selection of categorization parameters can be quantitatively guided by the FANOVA analysis discussed above. Though a detailed discussion is beyond the scope of this article, we note that, generally, those parameters on which the prediction parameters have significance 1st or 2nd order error dependence should be selected as categorization parameters. The two types of parameters as predictor variables are fed into a machine-learning algorithm (green box, detailed below) to make predictions. Suppose current time is t=0, the historical data for training the machine learning usually is available for a time period before t=0 (training period $T_{training}$). Specifically, the training data consist

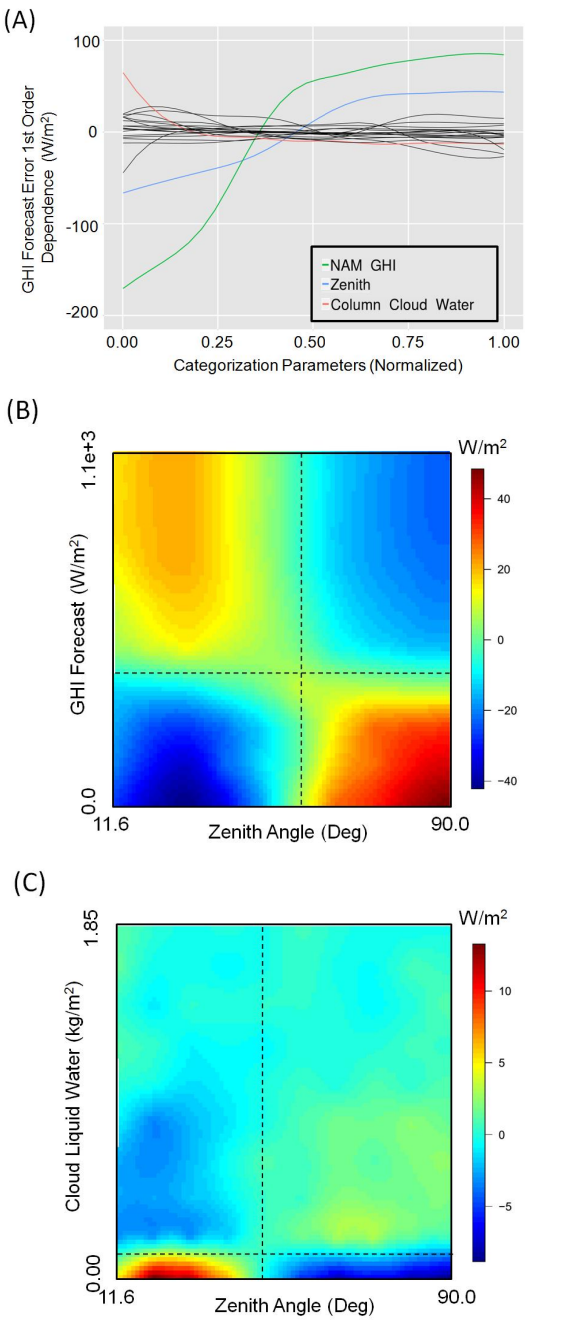

Figure 2. (A) North American Mesoscale Model (NAM) global horizontal irradiance (GHI) forecast error: 1st order dependence on input parameters derived using functional analysis of variance. Note that the three parameters on which GHI error depends upon most dominantly are GHI (green), zenith angle (blue), and column integrated cloud liquid water (red). The dependence on other parameters are shown in black. Note that in this plot all parameters are linearly scaled to [0,1] where 0 (and 1) is the minimum (and maximum) value encountered in the dataset so that 1st order dependences can be shown on the same plot. (B, C) Salient 2nd order GHI forecast error dependence are on GHI vs. zenith angle (panel B) and on column integrated cloud liquid water content vs. zenith angle (panel C).

of multiple physical models' forecasts for both the prediction parameters and the categorization parameters (both used as predictor variables) as well as measurements for the prediction parameters (used as response variables). Using such training data, a machine learning model is trained. By and large, a machine learning model is a statistical fitting between the predictions (of both prediction and categorization parameters) and the measurements (of prediction parameters) obtained by minimizing a certain cost function (purple box in Fig. 1). The cost function is typically the forecast accuracy metrics of interest to a forecast user such as RMSE or MAE, though more complex metrics may be needed depending on the use case of a forecast [32]. The trained machine learning model is then applied to the forecast period $T_{forecast}$ after t=0. The forecast for the prediction parameters and the categorization parameters from different physical models are used as the inputs to the machine learning model to forecast the prediction parameters in $T_{forecast}$.

From the test cases, the machine-learning based model blending approach proves to be quite generic. As long as an appropriate set of categorization parameters are selected, forecast error reduction with respect to the best individual input physical model can be achieved with an array of supervised machine learning algorithms which are from different underlying mathematical formulation. Tested machine learning algorithms include random forest (RF) [28,29], gradient boosted regression model (GBM) [33,34], recursive partitioning and regression tree (RPART) [35], support vector machine (SVM) [36,37], perceptron neural nets (NN). Unless stated otherwise, the results presented below are obtained using RF learning, although GBM, SVM, NN models provide comparable performance. As to be published elsewhere further improved forecast performance may be obtained by using a multiple-expert machine learning approach.

III. Result

A. Forecast Error Reduction w.r.t Best Individual Models

To investigate the forecast performance of the machine learning based model-blending approach we start with a case study in which three individual high resolution numerical weather prediction (NWP) models, NAM, RAP, and HRRR are blended. For validation of the forecasts, we use measurement data of four parameters, GHI, DNI, temperature at 2 m above ground (T2m), and wind speed at 10 m above ground (W10m). The choice of the four parameters is partially motivated for setting the stage for applying the same approach in solar power forecasting as these are the four input parameters for typical irradiance-to-power models [38,39]. GHI, T2m, W10m forecasts are taken from the NWPs directly, while DNI values are calculated from the vertical atmospheric and cloud profiles (temperature, pressure, humidity, cloudy liquid water and ice content) and surface albedo forecasted by the NWPs using a plane-parallel multi-layer radiative transfer model [40].

For the case study, daily 06z run of the NAM model and 11z run of the RAP and HRRR models are used to extract the forecast of GHI, DNI, T2m, and W10m. Forecast 5 to 21 hours ahead for NAM, 0 to 15 hours ahead for both RAP and HRRR) at the seven NOAA Surfrad stations [27] noted above are extracted and validated against the measurement.

Other than the four parameters for prediction, other categorization parameters extracted from the NWPs are cloud base and top heights, cloud column integrated ice content, cloud column integrated liquid water content, total cloud cover, ground pressure, ground relative humidity, surface rain as well as diffuse horizontal irradiance (DHI), solar azimuth angle, solar zenith angle calculated using the radiative transfer model. Pertaining to the result presented below, for the forecast of any given day, historical data of the prediction and categorization parameters of the 60 days immediately beforehand are used for training of an RF learning model. The validation time period is from 2013-07-05 to 2014-01-21.

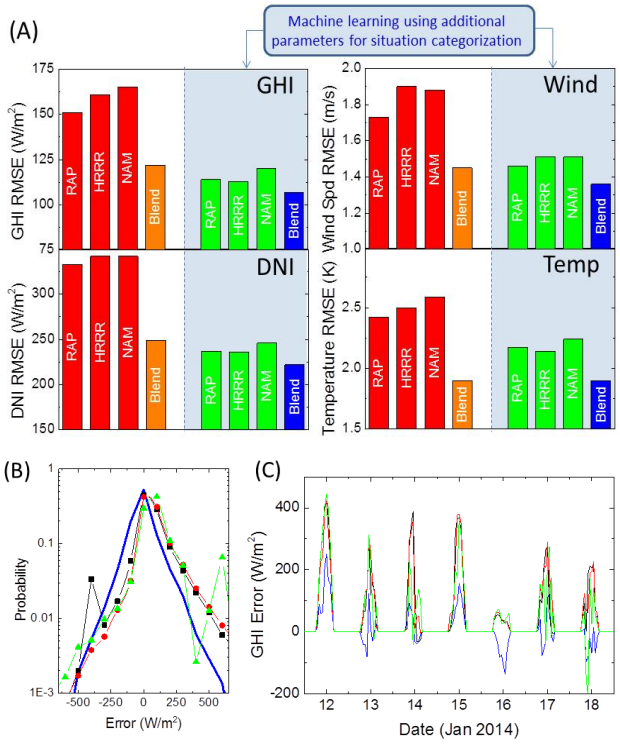

Figure 3 (A) The four panels summarize the forecast error of global horizontal irradiance (GHI), direct normal irradiance (DNI), temperature at 2 m above ground (T2m), and wind speed at 10 m above ground (W10m) using different methods. Red: uncorrected RAP, HRRR, NAM; orange: blending of RAP/HRRR/NAM without inputs of additional parameters for situation categorization; Green: ML correction of individual models with additional parameters for categorization; Blue: ML blending of three models using additional parameters for situation categorization. (B) Histogram of GHI forecast error distribution. (C) Comparison of GHI forecast error in an exemplary week for the Surfrad site at Pennsylvania State University (2014-01-12 to 2014-01-18). For (B) and (C) RAP error is shown in black, HRRR in red, NAM in green and the model blending in blue.

As a summary of the forecast result, the root mean square error (RMSE) of the four parameters predicted by machine learning based blending as well as extracted from the individual NAM, RAP, HRRR models is shown in Figure 3A. The RMSE of GHI forecast from uncorrected RAP, HRRR, NAM is respectively 151, 161, 165 W/m^2 (red bar). The RMSE of the model blended forecast based on machine-learning including parameters for categorization is 107 W/m^2 (blue bar), a ~ 30% improvement with respect to the best individual model. Similar degree of improvement is also seen for DNI, T2m, and W10m as shown in the other three panels of Fig.3A.

Fig 3B shows the histogram of GHI forecast error distribution. Compared to the three individual model (black, red, green lines), the blended model (blue) significantly reduced large positive forecast error by nearly one order of magnitude. Moreover as a typical example, a week of GHI forecast error from 2014-1-12 to 2014-1-18 at the Pennsylvania State University (PSU) Surfrad station is shown in Fig. 3C. Note that on January 12 and 15, even though the three individual RAP/HRRR/NAM models essentially have the same positively biased forecast, the model blending nevertheless substantially reduced the forecast error.

B. *Application in solar power forecasting.*

The machine-learning based model blending approach may also be applied towards PV power forecast. As a test case, the following result is for a 20.2MW PV site (multi-crystalline silicon module, single axis tracking in Marana, Arizona). Three NWP products are used as inputs for model blending: the daily 18z run of the NAM model (5 km resolution), the 15z run of the short range ensemble forecast model (SREF, Advance Research WRF central member on the 40 km resolution AWIPS grid 212) and the18z run of the GFS model (0.5 degree resolution). For all three models, the forecasts used are for the 24 hours of the next day (local time). Vertical atmospheric and cloud profiles at the site location are converted into GHI, DNI forecast using a plane-parallel multi-layer radiative transfer model. The GHI, DNI as well as the forecasted temperature at 2 m above ground and wind speed at 10 m above ground are used as inputs for an irradiance-to-power model adapted from [38] to calculate the forecast of the output PV AC power which is compared to the measured PV AC power.

For the validation period of 2014-01-23 to 2014-08-20, Fig.4A shows the 30 day rolling average of the MAE (excluding night time) for the model blending (red) vs. those derived from the three individual NWP models GFS (green), NAM (cyan), and SREF (purple). Clearly, the model blending consistently has the lowest MAE. For the validation period, the averaged MAE is 2.0 MW for model blending versus 2.9 MW for NAM, 2.9 MW for SREF, and 3.1 MW for GFS (Fig. 4B). Again model-blending enabled a ~30% error reduction compared to the best input model.

IV. Discussion

A. *The role of atmospheric state parameters for situation categorization in improving blended forecast accuracy.*

The fact that statistical correction and multi-model approaches may improve the accuracy of NWPs in itself is not that surprising, since unlike a general purpose NWP, model blending has to predict only a limited number of locations and atmospheric state parameters which has corresponding measurements for error correction. Here we focus on the role of including additional state parameters for situation categorization which may provide additional enhancement to conventional method exemplified by equally or unequally weighted ensemble averaging.

Turning back to the GHI forecast test case at the seven Surfrad stations, if we only feed the machine learning with the parameter to be predicted without categorization parameters, e.g. provide only GHI from RAP, HRRR, NAM for GHI prediction, the blended forecast has an RMSE of 122 W/m^2 (Fig.3A orange bar). In comparison, model blending including additional categorization parameters has an RMSE of 107 W/m^2 (Fig. 3A blue bar), an improvement of 12%. In fact, the simplistic model blending result is not as good as forecast based on correcting an individual model by including categorization parameters. If we use the prediction parameters and categorization parameters from an individual model (RAP, HRRR, or NAM) for training the machine learning model and for forecasting, such "corrected" RAP,

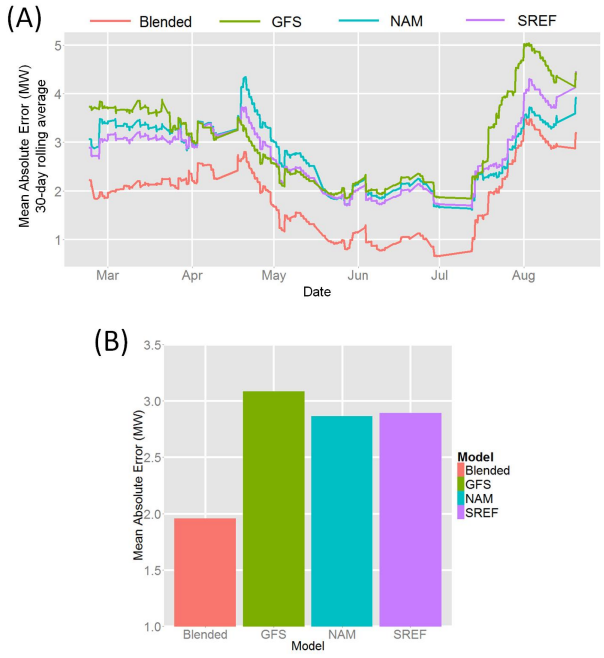

Figure 4.(A) Comparison of PV power forecast accuracy (MAE 30 day rolling average) derived from model-blending (red) and three individual NOAA NWP models GFS (green), NAM (cyan), and SREF (purple). (B) Bar plot showing average forecast MAE. Validation period is from 2014-01-23 to 2014-08-20.

HRRR, or NAM has a GHI forecast error of 114, 113, and 120 W/m^2 (Fig. 3A green bars). Similarly to the GHI case, the model blending including categorization parameters yields ~10% improved accuracy for DNI and wind compared to the cases without categorization parameters (Fig. 3A).

Generally speaking, the improvement offered by multi-model blending approach comes from canceling out the systematic bias error of individual forecast models. For conventional equally or unequally weighted ensemble averaging, the cancellation is for systematic error averaged over all possible atmospheric states. However, model error is not necessary equal in different sub-regions of the forecasted atmospheric states. For example, we have seen previously that the error of GHI forecast by NAM model is dependent on what is the combination of GHI (Fig. 2B) vs. zenith angle or column integrated cloud liquid water vs. zenith angle (Fig. 2C). Thus it is reasonable to postulate that improved blending results can be obtained if one takes into account localized difference in the error of individual models and provides appropriate adjustments depending on the parameters for categorization. This postulation is the core of this model-blending approach.

Indeed the advantage of model-blending with respect to simplistic model averaging can be illustrated when comparing the FANOVA results between the following two cases. As a reference, we calculate an ensemble model using an equal-weight linear combination of GHI forecast by the RAP, HRRR, and NAM models. In Fig. 5A the first order FANOVA error dependence of such averaged (labeled "Ensemble") GHI forecast is compared to forecast from model blending with categorization parameters (labeled "Model Blending"). The parameters for which FANOVA error dependence is calculated are listed in the insert box in Fig. 5A. All parameters have been linearly rescaled to [0,1] from their corresponding minimum and maximum values so that they can be overlaid on the same *x*-axis. Note that the ensemble averaged forecast (left panel) has significant 1st order error dependence - spanning from around -100 to +50 W/m^2 – on several parameters such as the HRRR model predicted GHI (yellow line). In comparison, the model blending result (Fig. 5A right panel) has much reduced first order error spanning from -30 to + 30 W/m^2. Similarly model blending reduces 2nd order error. Fig. 5B and 5C show the comparison of one of the dominant second order error dependences on zenith angle and GHI. Ensemble averaged forecast has large 2nd error ranging from around -15 W/m^2 (Fig. 5B blue region with small NAM GHI and small zenith angle) to around +15 W/m^2 (Fig. 5B red region with small NAM GHI forecast and large zenith angle.) In contrast, model blending forecast reduces the 2nd error to between around -5 to +5 W/m^2. Such reduction of error of 1st order and 2nd order, which is the basis for an overall improved forecast accuracy (Fig. 3 and 4), clearly cannot be obtained without involving the additional categorization parameters as inputs.

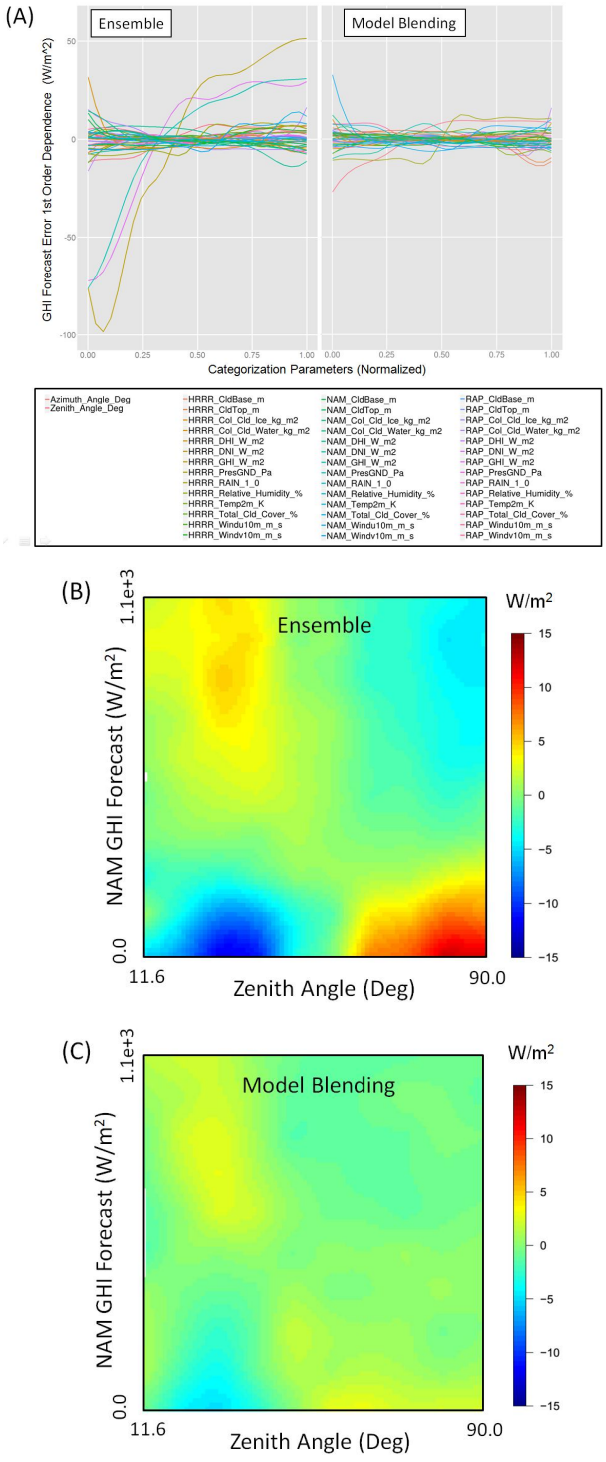

Figure 5. (A) First order GHI error dependence on all input parameters. The comparison is between ensemble model averaging (left) and model blending using categorization parameters (right). For plotting, all parameters are linearly scaled to [0,1] where 0 (and 1) is the minimum (maximum) value encountered in the dataset. Panels (B) and (C) show second order GHI forecast error dependence on NAM GHI and zenith angle. (B) is for forecast using ensemble model averaging and (C) is for model-blending with categorization parameters.

V. Conclusion

In conclusion, we have introduced a versatile machine-learning based multi-model blending approach for

enhancing the accuracy of weather forecasting in general and in particular for enhancing renewable energy forecast. From the test cases, we found the framework to be quite generic. Error reduction with respect to the best individual input model is achieved for different meteorological parameters as well as solar PV power. Validation results over an extended period of time at different locations in US show repeatedly over 30% improvement in solar irradiance/power forecast accuracy compared to forecast based on the best individual model.

The most salient feature of the forecast approach is to include - in addition to the parameter to be forecasted directly (such as GHI or PV power for solar power forecast) - parameters for weather situation categorization as inputs to the machine-learning. The motivation for doing so is the recognition that the forecast bias error of an individual physical model is "localized", i.e. dependent upon the categorization parameters. Such "localized" error dependence is quantified using functional analysis of variance. Compared to conventional ensemble approaches, which largely reduce mean errors of the individual models, the machine-learning based model-blending approach effectively reduces "localized" error of the individual models. We infer that such "localized" or situation dependent error correction underlies the overall improvement of the prediction accuracy brought by the machine-learning based model-blending approach.

ACKNOWLEDGMENT

This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represented that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

REFERENCES

[1] J.K. Lazo, M. Lawson, P.H. Larsen, and D. M. Waldman, "U.S. Economic Sensitivity to Weather Variability", *Bull. Amer. Meteor. Soc.* vol. 92, pp. 709-720, Jun. 2011.

[2] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, K. D. Brabandere, "Photovoltaic and Solar Forecasting: State of the Art," Technical report, IEA PVPS T14-01:2013, 2013.

[3] R. Perez, E. Lorenz, S. Pelland et al ,"Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe," *Solar Energy*, vol. 94, pp. 305-326, Aug. 2013.

[4] R. Marquez, C. F. M. Coimbra, "Intra-hour DNI forecasting based on cloud tracking image analysis". *Solar Energy* vol. 91, pp. 327 – 336, May 2013.

[5] Y. Chu, L. Nonnenmacher, R. H. Inman, Z. Liao, H. T. C. Pedro, C. F. M. Coimbra, "A smart image-based cloud detection system for intra-hour solar irradiance forecasts," *J. Atmos. Ocean. Technol.* vol. 31, pp.1995 – 2007, Sep. 2014.

[6] J. Kleissl, P. Mathiesen, C. Collier, "A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting", *Solar Energy* vol. 92, pp. 47–61, Jun. 2013

[7] C. Paoli, C. Voyant, M. Muselli, M. Nivet, "Forecasting of preprocessed daily solar radiation time series using neural networks." *Solar Energy* vol. 84, pp. 2146 – 2160, Dec. 2010.

[8] C. Chen, S. Duan, T. Cai, B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Solar Energy* vol. 85, pp. 2856 – 2870, Nov. 2011.

[9] R. Margolis,C. Coggeshall, J. Zuboy, "Integration of solar into the U.S. electric power system", Chap 6 in SunShot Vision Study, US Department of Energy, 2012. (http://energy.gov/eere/sunshot/sunshot-vision-study)

[10] K. Orwig, B.-M. Hodge, G. Brinkman, E. Ela, M. Milligan,V. Banunarayanan, S. Nasir, J. Freedman, "Economic evaluation of short-term wind power forecasts in ERCOT: Preliminary results." 11th Annual Int. Workshop on Large-Scale Integration of Wind Power into Power Systems as Well as on Transmission Networks for Offshore Wind Power Plants Conf., Lisbon, Portugal, 2012.

[11] EnerNex Corp. "Solar Integration Study for Public Service Company of Colorado. Denver, CO: Xcel Energy", 2009.

[12] EnerNex Corp, "Eastern Wind Integration and Transmission Study", NREL/SR-550-47078. Knoxville, TN. (Jan. 2010). http://www.nrel.gov/docs/fy10osti/47078.pdf

[13] J. Zhang, B.-M. Hodge, S. Lu, H. Hamann, B. Lehman, J. Simmons, E. Campos, V. Banunarayanan, "Baseline and Target Values for PV Forecasts: Towards Improved Solar Power Forecasting" Proceeding of 2015 IEEE Power & Energy Society General Meeting, In press.

[14] P. L. Houtekamer, H. L. Mitchell, "Data Assimilation Using an Ensemble Kalman Filter Technique". Mon. Wea. Rev., vol. 126, pp 796–811, Mar. 1998.

[15] P. L. Houtekamer, H. L. Mitchell. "A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation." *Mon. Wea. Rev.*, vol. 129, pp. 123–137, Jan. 2001.

[16] H. R. Glahn, D. A. Lowry, "The Use of Model Output Statistics (MOS) in Objective Weather Forecasting." *J. Appl. Meteor.*, vol. 11, pp. 1203–1211. Dec. 1972.

[17] R. L. Vislocky, J. M. Fritsch, "Improved Model Output Statistics Forecasts through Model Consensus." *Bull. Amer. Meteor. Soc.*, vol. 76, pp. 1157–1164, Jul. 1995.

[18] T. DelSole, X. Yang, M. K. Tippett, "Is unequal weighting significantly better than equal weighting for multi-model forecasting?" *Q. J. Roy. Meteor. Soc.* vol.139, pp. 176-183, Jan. 2013.

[19] T. N. Krishnamurti, C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, S. Surendran "Improved Weather and Seasonal Climate Forecasts from MultimodelSuperensemble, " *Sciences*, vol. 285, pp. 1548-1550, Sep. 1999.

[20] A. E. Raftery, T. Gneiting, F. Balabdaoui, M. Polakowski, "Using Bayesian Model Averaging to Calibrate Forecast Ensembles", *Mon. Wea. Rev.*vol 133, pp. 1155-1174 May 2005.

[21] P. Cudre-Mauroux, H. Kimura, K.-T. Lim, J. Rogers, R. Simakov, E. Soroush, P. Velikhov, D. L. Wang, M. Balazinska, J. Becla, D. DeWitt, B. Heath, D. Maier, S. Madden, J. Patel, M. Stonebraker, and S. Zdonik. "A demonstration of scidb: a science-oriented dbms."*Proc. VLDB Endow.* Vol. 2, pp. 1534–1537, Aug. 2009.

[22] Y. Chen, X. Xu, P. Li, S. Lu, S. Huang, W. Lu, K. Brown "GeoMix: Scalable Geoscientific Array Data Management", *Proceeding of the Industrial Track of 13th ACM/IFIP/USENIX International Middleware Conference*, Article No. 1, Dec 2013.

[23] T. Hastie, R.Tibshirani, J. Friedman, "The Elements of Statistical Learning Data Mining, Inference, and Prediction" 2nd Ed., *Springer-Verlag* 2011.

[24] G. Tesauro, D. Gondek, J. Lenchner, J. Fan and J. Prager, "Simulation, learning and optimization techniques in Watson's game strategies." *IBM J. Res. Dev. Vol.* 56, pp. 16:1-16:11 Jan. 2012.

[25] G. Hooker, "Generalized functional ANOVA diagnostics for high dimensional functions of dependent variables*", J. Comput. Graph. Stat.*, vol. 16, pp. 709-732, Sep. 2007.

[26] E. Rogers, G. DiMego, T. Black, M. Ek, B. Ferrier, G. Gayno, Z. Janjic, Y. Lin, M. Pyle, V. Wong, W-S. Wu, J. Carley, "The NCEP Northern American Mesoscale Modeling System: Recent Changes and Future

Plans", 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction, Omaha, NE, May 2009.
[27] J. A. Augustine, J. J. DeLuisi, C. N. Long, "Surfrad – A national surface radiation budget network for atmospheric research", *Bull. Amer. Meteor. Soc.*, vol. 81, pp. 2341–2357, Oct. 2000.
[28] N. Meinshausen, "Quantile Regression Forests", *J. Mach. Learn. Res.* vol. 7, pp 983-999, Jun. 2006.
[29] N. Meinshausen, "Quantile Regression Forest", http://cran.r-project.org/web/packages/quantregForest/quantregForest.pdf
[30] S. G. Benjamin, G. A. Grell, J. M. Brown, T. G. Smirnova, "Mesocale Weather Prediction with the RUC Hybrid Isentropic-Terrain-Following Coordinate Model", *Mon. Wea. Rev.* vol. 132, pp. 473-494, Feb. 2004.
[31] C. R. Alexander, S. S. Weygrandt, T. G. Smirnova, S. Benjanmin, P. Hofmann, E. P. James, D. A. Koch, "High resolution rapid refresh (HRRR): Recent enhancements and evaluation during the 2010 convective season", 25th conference on Server Local Storm, Denver, CO, Oct. 2010.
[32] J. Zhang, A. Florita, B.-M. Hodge, S. Lu, H. Hamann, V. Banunarayanan, A. M. Brockway, "A suite of metrics for assessing the performance of solar power forecasting", *Solar Energy* vol. 111, pp. 157-175, Jan. 2015.
[33] J.H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *Ann. Stat.* vol. 29, pp.1189-1232, Oct. 2001.
[34] G. Ridgeway, "Generalized Boosted Regression Models" http://cran.r-project.org/web/packages/gbm/gbm.pdf
[35] T. Therneau, B. Atkinson, B. Ripley, "Recursive Partitioning and Regression Trees" http://cran.r-roject.org/web/packages/rpart/rpart.pdf
[36] K. P. Bennett, C. Campbell, "Support vector machines: Hype or hallelujah?" ACM SIGKDD Explorations Newsletter, vol 2. pp.1-13 Dec. 2000. http://www.acm.org/sigs/sigkdd/explorations/issue2-2/bennett.pdf.
[37] C.-C. Chang, C.-J. Lin. "LIBSVM: a library for support vector machines." 2001 (http://www.csie.ntu.edu.tw/~cjlin/libsvm)
[38] D. L. King, W. E. Boyson, J.A. Kratochvill, "Photovoltaic array performance model", Sandia Report, SAND2004-3535, 2004.
[39] W. De Soto, "Improvement and Validation of a Model for Photovoltaic Array Performance", Master's Thesis, University of Wiscosin-Madison, Madison, WI, 2004.
[40] J. Key, A.J. Schweiger, "Tools for atmospheric radiative transfer: Streamer and FluxNet." *Comput.Geosci.* vol. 24, pp. 443-451 Jun. 1998.