Big Data and Machine Learning for Applied Weather Forecasts Forecasting Solar Power for Utility Operations

Conference Paper · December 2015

DOI: 10.1109/SSCI.2015.79

2 authors:

Sue Ellen Haupt
National Center for Atmospheric Research

Branko Kosovic
National Center for Atmospheric Research

Big Data and Machine Learning for Applied Weather Forecasts

Forecasting Solar Power for Utility Operations

Sue Ellen Haupt and Branko Kosovic
Weather Systems & Assessment Program, Research Applications Laboratory
National Center for Atmospheric Research
Boulder, CO USA
haupt@ucar.edu

***Abstract*—To blend growing amounts of renewable energy into utility grids requires accurate estimates of the power from those resources for both day ahead planning and real-time operations. This requires predicting the wind and solar resource on those timescales. Accurate prediction of these meteorological variables is a big data problem that requires a multitude of disparate data, multiple models that are each applicable to a specific time frame, and application of computational intelligence techniques to successfully blend all of the model and observational information in real-time and deliver it to the decision-makers at utilities and grid operators. Considering that the capacity of renewable energy continues to grow an additional challenge includes selecting and archiving data for continuous retraining of machine learning algorithms.**

Keywords—big data; weather forecasting; renewable energy; solar power forecasting

I. INTRODUCTION

Weather forecasting is one of the original computational challenges. From the time that L.F. Richardson in 1922 imagined a room of human "computers" numerically solving the primitive equations of fluid mechanics [1], meteorologists have been seeking to use big data and the best of numerical methods to improve forecasts. In fact the first problem on the first operational computer, the Electronic Numerical Integrator and Computer (ENIAC) at Aberdeen Proving Grounds in 1950 was a filtered version of these equations set up by Jules Charney, John von Neumann, and R. Fjortoft [2]. Thus was born meteorologists' love of computing and using and producing big data. Since then, the details included in the numerical models, as well as the spatial and temporal resolution employed have grown rapidly, continuously challenging computing capability. At the same time the field has been quick to employ advanced statistical and computational intelligence methods.

Current weather prediction efforts are an important real-time challenge and many applications rely on its accuracy, including aviation safety, defense planning, energy applications and beyond. To meet those expectations, numerical weather prediction (NWP) models are run hourly at high resolution and include tens of millions of grid cells, include physics packages that solve their own series of equations. These physics packages model or parameterize incoming and outgoing radiation, cloud physics, shallow and deep convection, boundary layer turbulence, land surface interaction with the fluid atmosphere, and more.

The nonlinear dissipative fluid equations are chaotic, implying sensitivity to initial conditions. This fact leads to two specific methods to deal with this chaos in modeling atmospheric flow: assimilation and ensemble prediction. Assimilation involves blending observed data into the initial model state. These data are seldom located on the grid and are often of disparate nature, leveraging attempts to sample the horizontal and vertical extent of the atmosphere and the land and sea surface boundaries. Ensemble modeling embraces the chaotic nature of the flow and seeks to provide multiple possible realizations of the development of the weather event. Some centers use upward of 50 model ensemble members to form a probability density function (pdf) of the development of the weather.

Finally, the best predictions blend as much data, models, and methods as possible via postprocessing to bring together information from multiple sources in order to improve the deterministic forecast as well as to quantify its uncertainty. Of course training these postprocessing methods requires a large amount of both model and observational data and will be discussed in more detail below. Some of the best methods involve rich data mining techniques that blend computational intelligence with knowledge of the physics and dynamics of the system. Such systems can be quite complex [3].

An example application that leverages all of these data issues is forecasting for variable renewable energy, namely wind and solar power. Accurate forecasting of wind and solar power is essential in order to reduce the levelized cost of renewable energy and effectively and efficiently integrate these

This work is funded by the US Department of Energy under Award Number DE-EE0006016.

variable energy sources into electric grid operations. Utilities and independent system operators (ISOs) rely on such forecasts to blend this variable yet valuable resource into the energy system. They need to allocate their resources a day or more ahead and correct predictions of the inexpensive solar and wind fuel allows minimizing the marginal cost of energy while assuring sufficient power to meet the load. In real-time, grid operators must have very short-range forecasts (referred to as nowcasts) to meet energy demand and minimize the cost of running excessive reserves. Thus modern meteorologists forecast the wind speed expected at turbines and irradiance at solar panels.

This paper lays out the issues for problems such as forecasting the weather for renewable energy. This is certainly a big data problem as discussed in Section 2. Section 3 provides a case study of the SunCast Solar Power Forecasting System designed by the National Center for Atmospheric Research (NCAR). The final section 4 discusses the issues and prospects for future applications.

II. Applied Weather Forecasting as a Big Data Challenge

A. Data Volume

Numerical Weather Prediction (NWP) is a big data issue, requiring some of the largest computations for real-time processing. NWP ingests large amounts of observational data for assimilating into the models and solves the nonlinear Navier-Stokes equations on grids upwards of 100 million grid cells with time steps on the order of 20 s. This implies that on the order of 18 billion of each model variable is handled hourly. Of course not all of that data is stored, with about 30 output variables being stored at 15 min increments and another 700 stored at hourly increments.

B. Data Variety

The data that are combined to provide a forecast typically include several different types of observations, some of which sample the surface variables at convenient locations that seldom correspond with the NWP grid points, while others depict the atmospheric vertical profile at particular locations. Yet other variables are sampled on a horizontal grid but with differing elevations, such as is true for satellite variables. Even where such observations are gridded, they may not be on the same type of map projection as are the NWP data. This all means that interpolation is a necessary step of any forecast process. Satellite irradiance values often depict the cloud top temperature. Since different clouds appear at differing atmospheric levels, these are indicative of cloud top height. Some satellite instruments, however, look through the clouds to determine properties of the atmosphere in vertical profiles. Yet other instruments, such as surface-based radars, scan the environment and provide reflectivity in terms of distance, angle, and azimuth of the beam. Thus, all of these differing types of data with differing types of grid systems must be coordinated in order to provide a sensible picture of the current weather before it is implemented into the forecasting system. The existence of standards and standardized formats for meteorological data, including metadata, significantly reduces the possibility of errors when processing disparate data. Unfortunately, no such standards exist at present for data collected at power plants. Standardization is essential for development of accurate, robust, and error free renewable energy forecasting systems.

One common standardization issue is the time stamp on the observations. Although all observations and NWP data should include a time stamp and that it should be in universal time, that is not always the case. In particular, it is common for specialized observations to be listed in local time. For such observations, it is sometimes confusing whether or not there is a time change when moving from standard to daylight savings time or vice versa.

Another issue is averaging time of observations. There are many small networks of specialized weather observations (mesonets). Although some of these are standardized, not all follow standardization on reporting details and averaging periods. For instance, one dataset that brings together observations from a variety of mesonets includes data with differing averaging periods, with hourly data including everything from averages of 10 Hz data for the full hour, for 15 min, for 5 min, for 1 min, and even for an instantaneous value. Some observations are recorded at the top of the hour, while others are an average of the prior hour. It is not a simple process to standardize the values that are provided. In using such data, it is important to read the information available on details of each type of observation and to be aware of best methods to deal with disparity.

C. Data Velocity

Having large amounts of data arrive on disparate time scales creates a huge challenge to processing. For a full system to operate, one must prepare for different times of arrival for each NWP model and each source of observation. This means that as data arrives, it must be matched to its valid time and steps taken to account for any lags before blending it with data from other models, observations, or systems.

D. Data Variablity

The realities of receiving the different data types in real-time implies that one must be prepared for any of the data sources to be delayed and have plans for graceful degradation of predictions when certain sources of data are not available in time to provide the real-time prediction. Because this is a frequent occurrence, fall-back routines are necessary for each type of data that may be missing. Because the machine learning algorithms are trained to optimize on having all of the data, it is necessary to also provide forecast model systems that assume missing data. Note that this process becomes even more complex when more than a single type of data is missing at the time that the forecast must be delivered.

E. Data Veracity

Data quality is a critical issue when training the computational intelligence models as well as in real-time. There are frequent issues with incorrect data that must be identified. For instance, there is a range of expected temperature values and when an observation is far from the expected range for the season and time of day at a location, it can be flagged for potential error. An additional check could be made on the previous value of temperature to see if the change in time is within reason. Note that for weather variables, however, one must take into account that occasionally rapid changes or anomalous values may be real. Temperature changes could occur rapidly upon frontal passage. There are occasional extreme values of each weather variable. For instance, during times of flooding, precipitation observations could appear anomalous when in fact, they are correct for that unusual case. This implies that quality control algorithms must be carefully constructed to identify these possibilities.

F. Data Complexity

These issues of data volume, variety, velocity, variability, and veracity point to the difficulties of attempting to blend these data to provide accurate forecasts in real-time.

For example, observations are used for at least three purposes in making the forecast: 1) training the computational intelligence algorithms, 2) identifying the current conditions to provide the necessary information for the current prediction, and 3) assimilation into the NWP models. After the prediction is made, data are again required for verification and validation. How to use the different sources of data well for each of these purposes without compromising the other uses is challenging.

III. A CASE STUDY: NCAR'S SUNCAST SOLAR POWER FORECASTING SYSTEM

An example of weather forecasting for a particular application is variable renewable energy wind and solar power, in particular. As the penetration of renewables increases, it becomes more difficult to merge it optimally with more conventional energy sources. Utilities and independent system operators (ISOs) are addressing the issue by employing forecasts of that energy. Of course, the wind and solar irradiance available depends on the weather. NCAR is working closely with utilities and ISOs to produce forecasts to meet their needs. Forecasts are necessary on very short and somewhat longer timescales. Grid operators require real-time situational awareness and best estimates for short-term changes to adapt the output of reserve units to accommodate more renewable generation or prepare for a rapid decrease in those sources. On a somewhat longer timescale, they make unit allocations a day ahead (or several days ahead for a weekend or holiday) that require knowledge of the output of the renewable resources.

Here, we focus on solar power forecasting. NCAR leads a Public-Private-Academic Partnership funded by the US Department of Energy (DOE) to advance solar power forecasting. To do this, one must be able to forecast the aerosols and clouds accurately. Cloud forecasting, in particular, has proven challenging in the past and this project seeks to advance the state-of-the-science to improve irradiance forecasting. To this end, NCAR has stood up a full solar power forecasting system, SunCast, comprised of both short-range (nowcast) components as well as longer range NWP models (see Fig. 1).

A. NWP Forecasts

The NWP forecasts for SunCast include models run by the National Center for Environmental Prediction (NCEP) of the National Oceanographic and Atmospheric Administration (NOAA), other national centers, and our own WRF-Solar.

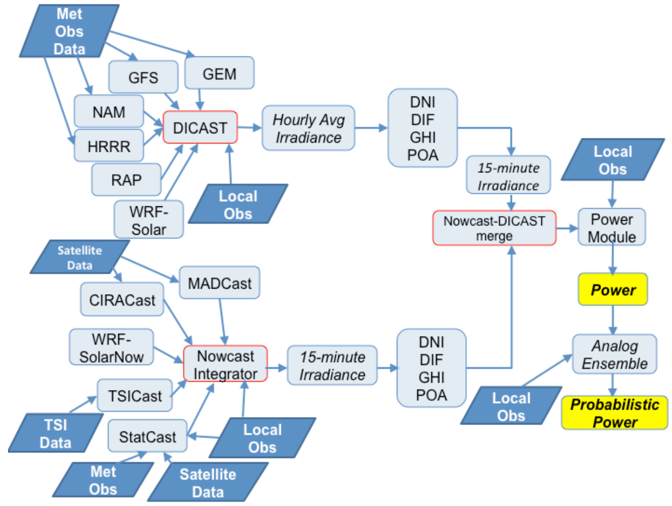

Figure 1. Diagram of data flow in the SunCast Solar Power Forecasting System.

These physics and statistically-based models are blended seamlessly and assigned optimal weights with the computational intelligence Dynamic Integrated Forecast System (DICast®) [4]. DICast is an automated forecast system designed to emulate the human forecast process. It examines current Numerical Weather Prediction (NWP) model data and generates forecasts based on empirical relationships developed from historical model data and observations. The multi-model solution is critical for reducing the forecast errors as the system. It uses a two step process: first it statistically corrects the bias dynamically of each input model in a process known as Dynamic Model Output Statistics (DMOS) [5], and second it optimally blends the models at each lead time. An advantage of DICast is that it is tuned to optimize the model blending using 90 days of data as compared to other methods that tend to require a year or more of data for training.

The current configuration of DICast used in SunCast uses seven NWP models (Fig. 2) and observations. The Global Forecast System (GFS), North American Model (NAM), Rapid Update model (RAP), and High Resolution Rapid Refresh (HRRR-NCEP) models are operational models run by NCEP, while the HRRR-ESRL is a research model developed at the Earth System Research Laboratory of NOAA and has shown excellent results in near term forecasting of precipitation and temperatures. The Global Environmental Mesoscale (GEM) model is run at Environment Canada. A new implementation of the Weather Research and Forecasting (WRF) model is being specifically designed to better predict solar irradiance: WRF-Solar [6]. Thus data from each of these models must be brought in and blended in real time to produce forecasts hourly.

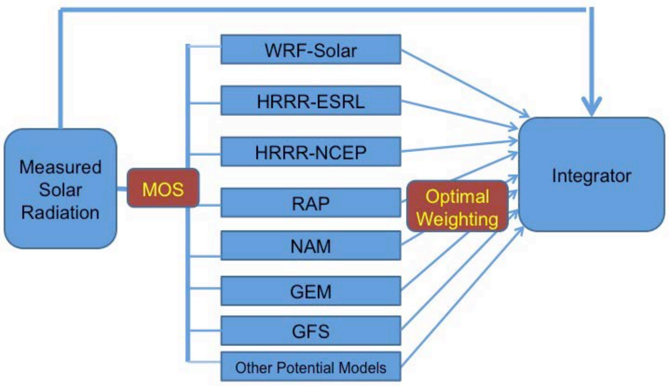

Figure 2. Diagram of the DICast blending process.

Each of these models has its own grid and timeframe as shown in Table 1 and described below.

1) HRRR

The High Resolution Rapid Refresh (HRRR) model is a relatively new National Center for Environmental Prediction (NCEP) model that covers the continental United States using a 3 km grid cell size and produces hourly 15 hour ahead forecasts every hour. A single forecast consisting of three-dimensional fields amounts to 365 Mb, while corresponding two-dimensional fields representing surface conditions account for additional 84 Mb for a total of approximately 450 Mb.

2) RAP

The Rapid Refresh (RAP) model version 2 preceded HRRR. It provided hourly forecast out to 18 hours every hour. Each RAP output file amounts to 55-60 Mb.

3) GFS as a typical global model

The Global Forecast System (GFS) is NCEP's global model. GFS is run at 2.5, 1.0, and 0.5 degrees globally. The model forecasts are produced every 6 hours out to 384 hours. Recently additional 0.25 degree simulation was added out to 168 hours. In our solar irradiance forecasting system 0.5 degree forecast is used 65-70 Mb.

4) Specialized runs such as WRF-Solar

WRF-Solar is a new branch of the WRF model designed specifically for improved forecasting of solar irradiance. This version includes an improved radiative transfer scheme, new shallow cumulus scheme, and output tailored to the specific application. Initial and boundary conditions for the WRF-Solar forecasting system are provided by the RAP analysis. The irradiances (global horizontal irradiance - GHI, direct normal irradiance - DNI, and diffuse irradiance - DIF) are output every iteration, or every 20 seconds and one-minute averages are computed.

Table 1. Details of several NWP model daily outputs.

Model	Forecast frequency	Hours ahead	Grid cell size	Daily output size [GB]
HRRR	hourly	15	3 km	130
NAM	hourly	18	9 km	5.7
GFS	6 hours	384	0.5 degrees	68
GEM	12 hours	240	1 degree	4.3
WRF-Solar	irradiances only: 20 sec	30	3 km (CONUS) 1 km (2 subdomains)	4.2

Table 2. Details of several nowcast model daily outputs.

Model	Forecast frequency	Hours ahead	Daily output size [MB]
MADCast	15 minutes	6	2100
CIRACast	15 minutes	6	1.4
StatCast	15 minutes	3	13

B. Nowcast System

The Nowcast system is comprised of five models, each with its own sweet spot for producing a most accurate forecast (Table 2).

1. WRF-Solar is run in Nowcasting mode at a coarser resolution over CONUS in order to produce hourly output. It predicts out to 6 hours with approximately a 1 hr latency to complete the run.

2. The Multi-sensor Advection Diffusion foreCast (MADCast) system uses a the Multivariate Minimum Residual (MMR) scheme proposed by Aulighé [7, 8] to assimilate satellite infrared radiance observations into the dynamic core of WRF, which then advects those observed clouds according to WRF dynamics. It also predicts out to 6 hours, but the latency is only about 10 min due to not employing the computationally expensive physics packages of WRF.

3. CIRACast, designed by Colorado State University's Cooperative Institute for Research in the Atmosphere (CIRA), leverages satellite-observed clouds and advects them with derived motion vector and model winds to estimate cloud coverage over the coming hours. Its latency depends on the time to process and ingest remotely the satellite and model wind data, typically around 15-30 min.

4. StatCast is a regime-dependent neural network model that ingests surface irradiance measurements, nearby weather data, and satellite data to estimate the clearness index (the observed surface irradiance divided by that available at the top of the atmosphere at that location) for the following 3 hours. It runs in a matter of seconds.

5. TSICast is designed by Brookhaven National Laboratory to observe current cloud cover with three total sky imager (TSI) cameras and deduce the height, base, location of the clouds in its line of sight, as well as the speed and direction of each cloud layer to predict where they will be in the next 15-

30 min [9]. TSICast takes 2-3 min to fully process the data and provide a very short range prediction.

These nowcasting methods leverage a variety of disparate observational data, statistical and computational intelligence methods, and physical understanding of the atmosphere to produce a "best practices" blended forecast. The amount of data produced by the Nowcast system is provided in Table 2. The data amount is directly proportional to the number of sites for which the forecast is produced. The Nowcast system is currently optimized via an expert system, but dynamic methods will be applied in the future. Each of these models has been shown to provide value in the system.

Observations used in the Nowcast system include irradiance, air temperature, and power output. The total amount of data received daily for all the sites for which forecasts is produced (14 Sacramento Municipal Utility District, 2 Xcel Energy, 9 SoCal Edison, and 25 Brookhaven National Laboratory) is approximately 35 MB. While the amount of data is modest, data quality and disparate data formats represent a processing challenge.

C. *Completing the Forecast*

The irradiance forecasts of the DICast and Nowcast system are combined and blended in the transition times (2 hrs - 6 hrs) to produce irradiance forecasts for every 15 min out to 3 hours then hourly out to 168 hours. The forecast variables are global horizontal irradiance (GHI), which is most useful for photovoltaic panel operations; Direct Normal Irradiance (DNI), which is the only component useful for concentrated solar plants; and Diffuse Irradiance (DIF) that relates the two.

Utilities are not really interested in the meteorological irradiance values, but instead wish to have a power prediction. The SunCast system trains a model regression tree (Cubist) on the relationship between the measured irradiance value and the power measurement. Cubist requires historical data for training and testing, then applies the empirical relationship in real-time to the irradiance forecast to produce a power forecast. The power conversion algorithm is trained for each solar site to be predicted. Although the training/testing procedure requires data, once trained, the algorithm runs in a matter of seconds.

Finally many of the utilities involved in the partnership wish to have probabilistic forecasts. Although NCAR could choose to run ensembles or to postprocess the various methods to estimate a pdf, we choose the Analog Ensemble (AnEn) approach. The Analog Ensemble operates on the premise that if a forecast was made in the past under certain meteorological conditions analogous to today's forecast, then the error in that historical forecast is likely to be similar to the error in today's forecast. So if we can identify such analogs in those past forecasts, that enables us to do two things: 1) correct today's forecast by the prior observed error, improving upon the best deterministic forecast, and 2) use the pdf of multiple analogs to provide an estimate of the uncertainty. This is a flow dependent uncertainty and has been shown to reproduce the forecast and its statistical reliably as well as or better than some of the best full ensembles of runs employed at the operational centers [10,11].

NCAR produces such probabilistic solar power forecasts in quasi-operational mode hourly and makes them available to the utility and ISO partners in the project. We are in the midst of a full assessment. Preliminary results indicate that each component improves upon baseline forecasts and that the SunCast system produces forecasts that are useable by the utility and ISO partners in integrating their solar resources into their energy mix. Evaluation plans include production cost modeling, which will estimate the monetary value of the forecasts and reserve analysis, which will study the changes in usage of the energy reserve units due to providing more confidence in the amount of solar energy available at each future time in the forecast.

IV. Summary and Looking Ahead

Because renewable energy is becoming more prevalent, it comprises a higher percentage of the energy capacity. Thus, one needs to be able to forecast its expected value, variability, and uncertainty. To do that requires blending NWP data from multiple centers, specialized models tuned to the location, and data observations from that location. Observations are also necessary to build computational intelligence algorithms that optimize the forecast. This application necessitates handling data that is large in volume, using it in real-time at high velocity, of high variety, and may be of questionable veracity. The complexity of blending all of this information to provide a forecast in time to be useful is a complex big data problem. Here we have described one such system developed by NCAR and its collaborators. The issues involved in making the forecast are numerous. As we work to make such forecasts ever more accurate, various issues must be addressed.

Modelers are always working to improve their forecasts. As computers become larger, we attempt to increase the resolution of the models. At some points, we hit limitations to our model resolution. For instance, as NWP models go to higher and higher resolution, we find that we are resolving processes that were previously parameterized. We are also modeling scales of the atmosphere that are different than where the model was constructed, and so the important physics may be a bit different. Although the NWP models are constructed to resolve large scale flow over large domains, large eddy simulations (LES) are appropriate for much finer scales. There is a "terra incognita" between about 1000 m and 100 m where the physics changes sufficiently that blending through those scales has been shown to be inappropriate [12].

We have seen that meteorologists run multiple realizations of the NWP models in recognition of the chaotic nature of the flow and sensitivity to initial conditions. As computers become bigger and faster, there is a tendency to want to run more ensemble members to better fill out the pdf of the forecast in order to make better predictions. It will be interesting to watch whether that approach is best, or whether the statistical learning methods that postprocess ensembles to fill out the pdf statistically are shown to compete well with large ensembles. The limiting case, of course, is exemplified in the analog ensemble method described above.

The data mining techniques that blend the models and the observations are also becoming more complex. For instance, deep machine learning is demonstrating some success in predictions in many fields. But such methods require yet more data. Again, it will be interesting to watch the balance between data volume and its smart usage develop over time.

Machine learning techniques and analog ensembles critically depend on archived data for training and sampling purposes. Considering the volume of model output and observation data sets, a subset of essential variables must be selected and archived at appropriate frequency for future use.

Accurate renewable power forecasting critically depends on the amount and quality of various types of data. In addition to power production, the data regularly collected at power plant include atmospheric variables: g GHI or DNI, wind speed and direction, and temperature, as well as pressure and humidity. In order to effectively use the collected data, the dataset format should be standardized. Particular attention should be given to comprehensive meta data including information about the instruments used to collect the data, accurate location information where the data were collected, and the time when the data were collected as well as instrument maintenance record. The location should be given in latitude and longitude based on the World Geodetic System (WGS) standard from 1984 which was revised in 2004 and the height above the surface. To avoid any confusion about time zones and seasonal time changes due to the "Daylight Saving Time," the time when the data was collected should be reported in internationally accepted Coordinated Universal Time or UTC.

The data should be organized and stored following one of the established portable data formats for ASCII data such as for example Met_Point or little_r formats (http://www.dtcenter.org/met/users/docs/presentations/MET_Tutorial_20090204/04_DataFormats.pdf)or self-describing binary formats such as GRIdded Binary or General Regularly-distributed Information in Binary (GRIB) format, Common Data Format (CDF, http://cdf.gsfc.nasa.gov/), Network Common Data Format (NetCDF, http://www.unidata.ucar.edu/software/netcdf/), or Hierarchical Data Format (HDF, http://www.hdfgroup.org/). Using well defined, documented, and widely used data formats significantly enhances utility of data sets and simplifies their processing and quality control.

Finally, such applications are moving toward cloud computing frameworks, which introduce their own complexities. The issues of disparate data arriving at different times and requiring blending to provide real-time forecasts will be complicated yet further as we wish to deploy on a larger variety of architectures.

Although these challenges are far from trivial, the prospects for application are quite promising. The need for improved forecasts of the renewable energy variables certainly are becoming increasingly important as we seek to blend more wind and solar energy into the grid. As the need for these forecasts increases, so will the solutions to the challenges.

ACKNOWLEDGMENT

The authors thank the entire SunCast team at NCAR and beyond. Special thanks go to Seth Linden, who collected information on data usage in the SunCast system.

REFERENCES

[1] L.F. Richardson, Weather Prediction by Numerical Process, Cambridge University Press, Cambridge, 1922.

[2] J.G. Charney, R. Fjortoft, and J. von Neumann, "Numerical integration of the Barotropic Vorticity Equation," Tellus, vol. 2, pp. 237-254, 1950.

[3] W.P. Mahoney, K. Parks, G. Wiener, Y. Liu, B. Myers, J. Sun, L. Delle Monache, D. Johnson, T. Hopson, and S.E. Haupt, "A Wind Power Forecasting System to Optimize Grid Integration," *IEEE Transactions on Sustainable Energy*, 3 (4), 2012, pp. 670-682.

[4] W. Myers, G. Wiener, S. Linden, and S. E. Haupt, "A consensus forecasting approach for improved turbine hub height wind speed predictions,"in *Proc. WindPower 2011*, Anaheim, CA, May 24, 2011.

[5] H. R. Glahn and D. A. Lowry, " The use of model output statistics (MOS) in objective weather forecasting," J. Appl. Meteor., 11**,** 1972, pp. 1203–1211.

[6] P. Jimenez-Munoz, J.P. Hacker, J. Dudhia, S.E. Haupt, J.A. Ruiz-Arias, C.A. Gueymard, G. Thompson, T. Eidhammer, and A.J. Deng, "WRF-Solar: An Augmented NWP Model for Solar Power Prediction," submitted to Bull. Amer. Met. Soc. 2015.

[7] T. Auligné, T., "Multivariate Minimum Residual Method for Cloud Retrieval. Part I: Theoretical Aspects and Simulated Observations Experiments," Mon. Wea. Rev., 142, 2014, pp. 4383-4398.

[8] T. Auligné, T., "Multivariate Minimum Residual Method for Cloud Retrieval. Part II: Real Observations Experiments," Mon. Wea. Rev., 142, 2014, pp. 4399-4415.

[9] Z. Peng, D. Yu, D. Huang, J. Heiser, S. Yoo, and P. Kalb, "3D cloud detecting and tracking system for solar forecast using multiple sky imagers," *Solar Energy*, 118, 2015, pp.496-519.

[10] L. Delle Monache, T. Eckel, D. Rife, B. Nagarajan, K. Searight, "Probabilistic weather predictions with an analog ensemble," *Mon. Wea. Rev*., **141**, 2013, pp. 3489-3516.

[11] S. Alessandrini, L. Delle Monache, S. Sperati, and J. Nissen, "Probabilistic power prediction with an analog ensemble," Submitted to *Renewable Energy*, 2015.

[12] J. C. Wyngaard, "Toward numerical modeling in the 'Terra Incognita'," J Atmos Sci 61, 20014, pp. 1816–1826