

1 **Solar Energy Prediction: An International Contest to Initiate**
2 **Interdisciplinary Research on Compelling Meteorological**
3 **Problems**

4 **AMY MCGOVERN ***

School of Computer Science, University of Oklahoma, Norman, Oklahoma

5 **DAVID JOHN GAGNE II**

School of Meteorology, University of Oklahoma, Norman, Oklahoma

6 **LUCAS EUSTAQUIO**

Niddel

7 **GILBERTO TITERICZ JUNIOR**

Petrobras

8 **BENJAMIN LAZORTHES**

Blia Solutions, France

9 **OWEN ZHANG**

zhonghua.zhang2006@gmail.com

10 **GILLES LOUPPE**

Department of EE&CS, University of Liege, Belgium

11

PETER PRETTENHOFER

DataRobot, peter@datarobot.com

12

JEFFREY BASARA

School of Meteorology, University of Oklahoma, Norman, Oklahoma

13

THOMAS M. HAMILL

NOAA/ESRL Physical Sciences Division

14

DAVID MARGOLIN

EarthRisk Technologies and Solutions

*Corresponding author address: Amy McGovern 110 W. Boyd St., Norman, OK 73019

E-mail: amcgovern@ou.edu

16 As meteorological observing systems and models grow in complexity and number, the size of
17 the data becomes overwhelming for humans to analyze using traditional techniques. Com-
18 puter scientists, and specifically machine learning and data mining researchers, are develop-
19 ing frameworks for analyzing big data. The AMS Committee on Artificial Intelligence and
20 its Applications to Environmental Science aims to bring AI researchers and environmental
21 scientists together to increase the synergy between the two. The AI committee has spon-
22 sored 4 previous contests on a variety of meteorological problems including wind energy,
23 storm classification, winter hydrometeor classification, and air pollution, with the goal of
24 bringing together the two fields of research. Although these were successful, the audience
25 was limited to existing environmental science researchers (usually 10-20 teams of people
26 primarily within the AMS community). For the 2013/14 contest, we expanded to a global
27 audience by focusing on the compelling problem of solar energy prediction and by having
28 the established forum Kaggle host our contest. Using this forum, we had over 160 teams
29 from all around the world participate. Improved solar energy forecasting is a necessary com-
30 ponent of making solar energy a viable alternative power source. This paper summarizes
31 our experiences in the 2013/14 contest, discusses the data in detail, and presents the win-
32 ning prediction methods. The contest data come from the NOAA/ESRL Global Ensemble
33 Forecasting System Reforecast Version 2 and the Oklahoma Mesonet with sponsorship from
34 EarthRisk Technologies. All winning methods utilized gradient boosted regression trees but
35 differed in parameter choices and interpolation methods.

1. Introduction

The increasing size and complexity of meteorological observational data and model output has demanded more investigation into how best to utilize this wealth of data. The challenge of analyzing large data volumes is not unique to meteorology. Computer scientists, and specifically machine learning and data mining researchers, are developing frameworks for analyzing big data for a range of applications. The AMS Committee on Artificial Intelligence and Its Applications to Environmental Science aims to bring AI researchers and environmental scientists together to increase the synergy between the two fields. The AI Committee has sponsored 4 previous contests on a variety of meteorological problems including wind energy, air pollution, winter hydrometeor classification, and storm classification (Lakshmanan et al. 2010) with the goal of bringing together the two fields of research to discuss a common challenge from multiple perspectives. The winners of the past contests presented in a special session at the AMS Annual Meeting that featured both the results and discussions of the various techniques used as well as how they could be applied to similar problems. While the discussions had been fruitful and attracted people from many different backgrounds, participation in the contests declined from year to year. For this past year's contest, we made significant changes to the contest format in order to increase participation and reach a much wider audience.

Our goal for 2013/14 contest was to determine which approach produces the best total daily solar energy forecast. We changed three key features of the contest organization. First, we used the year prior to the contest to gather and format a larger and more complex dataset for predictions. Second, we hosted the contest website on Kaggle, a popular platform for AI contests with a worldwide audience. Third, we extended the time window of the contest from just the fall to July through November and allowed contestants to submit and evaluate entries every day throughout the period. These changes resulted in an order of magnitude increase in the number of participants and a broadening of the participant pool from those in the existing meteorological community to scientists and engineers around the world.

2. Data

The forecast data used in this study came from the second-generation NCEP Global Ensemble Forecast System (GEFS) reforecast set described in Hamill et al. (2013). This data consists of an 11-member global ensemble initialized at 00 UTC every day from 1985 to present. Forecasts extend to +16 days lead time. The modeling system closely replicates the GEFS as it was implemented in 2012. The initial conditions for most of the data set used the Climate Forecast System Reanalysis (CFSR; Saha and Coauthors 2010) for the control initial condition and the ensemble transform with rescaling (Wei et al. 2008) for generating perturbed initial conditions. Forecast data was archived every 3 h to +72 h lead time, and every 6 h thereafter. More details are available in Hamill et al. (2013).

The Oklahoma Mesonet is a permanent mesoscale surface observing network of 120 remote, meteorological stations across Oklahoma (Brock et al. 1995; McPherson et al. 2007). The Mesonet represents a partnership of Oklahoma State University and the University of Oklahoma and managed by the Oklahoma Climatological Survey (OCS). Each station measures more than 20 environmental variables, including wind at 2 and 10 m, air temperature at 1.5 and 9 m, relative humidity, rainfall, pressure, solar radiation, and soil temperature and moisture at various depths. All sensors are mounted on or near a 10-m tower supported by three guy wires and powered via solar energy.

Downwelling, global solar radiation is measured by the LI-COR LI-200 pyranometer mounted on a boom that extends southward from the tower. Even so, measurements of solar radiation during early morning and late afternoon evening may be sensitive to obstructions to the east and west of the station. All solar radiation data are collected and transmitted to a central point every 5 minutes where (1) sensor-specific calibration coefficients are applied and (2) the data are quality controlled via automated algorithms and human inspection prior to distribution and archival (Shafer et al. 2000; McPherson et al. 2007).

The locations of the GEFS and Mesonet stations are shown in Fig. 1. Due to the coarseness of the GEFS grid relative to the Mesonet station spacing, contestants were pro-

vided with additional grid points well outside the Oklahoma state boundaries so that any interpolation techniques would not experience any interference from edge conditions.

3. Contest Setup

The contest was hosted by Kaggle, a company that developed a platform for hosting data mining competitions in addition to providing modeling support for a variety of Fortune 500 companies. For each competition hosted on the site, Kaggle provides pages for describing the competition and the rules, downloading the data, displaying real-time rankings of the participants, and discussions about the contests. The site also automatically manages submission of contestant entries and evaluation of the predictions. The continuous stream of contests on Kaggle has led to the development of a large community of contest participants who come from a wide range of backgrounds and from around the world. For these services and for access to its large user community, Kaggle charges a fee to companies who wish to host their contest through the site, but Kaggle also hosts research competitions for smaller contests organized by academic groups for no fee. EarthRisk Technologies sponsored the contest and provided the prize money.

For this contest, a small spatial subset of the 11-member ensemble data was extracted over Oklahoma and surrounding regions, consisting of forecasts at the +12, +15, +18, +21, and +24-hour lead times. To be coincident with the observational data, the reforecast data was extracted only back to 1994. These pervaded the forecast training data for the contest's 1-day solar-energy predictions. The forecast variables saved were mean sea-level pressure, skin and 2-meter temperature, 2-meter specific humidity, daily maximum and minimum 2-meter temperature, total precipitation in the last 3 hours, total column precipitable water, total column integrated condensate, total cloud cover, downward and upward short- and long-wave radiation flux at the surface, and upward long-wave radiation flux at the top of the atmosphere.

The data were split into training, public testing, and private testing sets. The training set timeframe extended from 1 January 1994 to 31 December 2007; the public testing set ranged from 1 January 2008 to 31 December 2009, and the private testing set ranged from 1 January 2010 to 30 November 2012. Teams could evaluate their predictions on the public testing set up to 5 times per day and optimize their algorithm based on the evaluation score. The final ranking of the teams was determined from the private testing set results, and the scores were not revealed until the contest concluded. Mean absolute error over all stations and days was chosen as the evaluation metric because it does not penalize extreme forecasts as greatly as root mean squared error.

In addition to the contest data, participants also received the results and source code for three benchmark methods that indicated how random selection and interpolation methods would perform on the dataset. The random normal benchmark input random numbers sampled from a normal distribution with a mean of 16 MJ m^{-2} and a standard deviation of 8 MJ m^{-2} . The other two benchmarks interpolated the GEFS mean total daily incoming solar radiation to the Mesonet sites using nonlinear approaches. One method fit a set of scaled Gaussian mixture models to the GEFS data with an expectation-maximization iterative approach similar to the method of Lakshmanan and Kain (2010). It produced a smoothed field that could be evaluated at any point in the domain and had an MAE of 4.02 MJ m^{-2} . The second approach was to use Catmull-Rom cubic splines to interpolate the nearest four grid points to each Mesonet site. The splines performed significantly better than the Gaussian mixture model approach with an MAE of 2.61 MJ m^{-2} although they did have a tendency to have larger extremes than the observed data. Once the spline code was provided to the contestants, 118 of the 160 teams were able to either equal or improve on its performance.

4. Gradient Boosted Regression Trees

One of the surprising outcomes of the contest was that all of the winning methods made use of the same regression technique, Gradient Boosted Regression Trees (GBRT) (Friedman 2001). GBRT robustly models the (volatile) daily solar energy output from spatio-temporal input variables. For this data, GBRT proved to be an accurate and effective off-the-shelf regression technique because i) it natively handles data of mixed type, ii) it is robust to outliers (through robust loss functions) and iii) it is non-parametric and has high predictive power.

Mathematically, GBRT is a generalization of boosting (Freund and Schapire 1995) to arbitrary differentiable loss functions L . The method considers additive models of the form

$$F_m(x) = \sum_{m=1}^M \gamma_m h_m(x) \quad (1)$$

where $h_m(x)$ are basis functions called *weak learners*. In GBRT, weak learners are regression trees (Breiman et al. 1984) that are learned sequentially using a forward stagewise procedure. More specifically, at each stage, $h_m(x)$ is chosen to minimize the loss function L via steepest descent (using the negative gradient of L at the current model F_{m-1}) while the step length γ_m is chosen using line search.

5. 1st place: Eustaquio and Titericz

The winning approach creatively combined the predictions from models that focused on different aspects of the input data as well as information about the spatial and temporal variability. At each Mesonet site, 13 GBRT models were trained. The first 11 models used input data from each GEFS ensemble member, and the other two used the medians and maximums of the GEFS variable values over all ensemble members. The models trained on each member incorporated data from the 4 GEFS grid points that surrounded each Mesonet site. The 5 intraday values for all 15 input weather variables were used from the 4 nearest

grid points, resulting in 300 input values per day. Additional descriptive variables (latitude and longitude from the GEFS and Mesonet, the station ID, and the distances between the Mesonet site and GEFS points) were also included. The aggregated models were trained on either the median or the maximum value of the ensemble variables and on the sum of the intraday values. All of the models were trained and optimized with 3 fold continuous cross-validation over consecutive 4 year periods. The Python implementation of the GBRT was used.

Once the individual models had been trained and once each produced solar energy predictions over the training time period, two optimized weighted ensembles were produced to create a consensus solar energy prediction for each site. The forecasts for each station were combined using the Nelder and Mead (1965) non-linear optimization algorithm to minimize the MAE of the consensus prediction. A second optimized ensemble was created by optimally weighting the predictions at nearby Mesonet sites to match the predictions at a particular site. The two weighted ensemble predictions were then simply averaged and then multiplied by 1.01 as a final bias correction. All of the models took 12 hours to run and resulted in an error of 2.11 MJ m^{-2} . For comparison, the mean daily production of all Mesonet sites was 16.7 MJ m^{-2} , resulting in a mean global error of 13%. It should be noted that no manual feature engineering was performed; the GBRT and the optimization routines did all of the feature selection and distance weighting on its own.

6. 2nd place: Lazorthes

As is often the case in predictive analytics, data preparation was the most important step in this project. Since the localization of the mesonet stations did not coincide exactly with the position of the GEFS nodes (see Figure 2), some transformations were necessary in the training and testing datasets. For each of the 98 Mesonet stations, a linear interpolation of the four nearest GEFS points (weighted by the distance) was carried out using the following

formula.

$$V_{\text{Mesonet}} = \frac{\sum_{i=1}^4 w_i V_{\text{GEFS}i}}{\sum_{i=1}^4 w_i} \quad (2)$$

$w_i = \max(0, 1 - d_i)$ and d_i was the Euclidian distances from the mesonet to the near GEFS nodes (assuming that the smallest distance between 2 GEFS is equal to 1).

Fifteen meteorological variables forecasted each day at 00 UTC for 5 different hours (at 12, 15, 18, 21, and 00 UTC the following day) were provided. We used these 75 weather features without any prior selection. Additional features were created by spatially or temporally averaging the original 75 weather variables. In addition, the elevation, the latitude and the longitude of the Mesonet stations and the month of the observation. At the end, 128 explanatory variables were defined.

All the data from the 98 Mesonet stations were gathered to obtain a single training set, a single testing set, and finally a single model for all stations. Some trials have been performed with separated datasets for each station but they never gave more accurate predictions. Consequently, the training dataset had 501,074 rows and the testing dataset had 176,068 rows.

The best accuracy was achieved with GBRT, using the implementation directly available in R (gbm package) with the mean absolute error (MAE). Random Forests (Breiman 2001) were also evaluated but were not retained because they were less accurate.

For each of the boosted trees, we used the following training settings: Mean Absolute Error (distribution = “laplace”), number of expansions between 2000 and 3000 (n.trees = 2000 or 3000), depth of the trees between 6 or 8 (interaction.depth = 6, 7 or 8), a learning rate of 0.05 (shrinkage = 0.05), an out of the bag proportion: 30% (bag.fraction = 0.7). An ensemble of 12 distinct gradient boosted regression trees improved the accuracy by reducing the risks of overfitting.

The mean absolute error of the 2nd place model was 2,128,116 J/m², as evaluated on the private test set. Knowing that the average daily incoming solar energy of the stations in the Mesonet is around 16,500,000 J/m², it therefore corresponds to a mean absolute error

213 of about 12.8%.

214 Some variables clearly appeared to be particularly important: the downward short-wave
215 radiative flux average at the surface (dswrf) and the precipitable water over the entire depth
216 of the atmosphere (pwat). Even if the other variables are less influential, they contribute to
217 improve the global accuracy of the model. Table 1 gives the top 10 most important variables
218 as ranked by their perfect of influence.

219 Fig. 3 is a 3D graphical representation of the model. The shape of the curve is typical
220 of models obtained by combining several regression trees and shows the dependence of the
221 model on incoming solar radiation and precipitable water and how the two terms interact.
222 As physically expected, increased precipitable water results in lower observed solar energy
223 for a given amount of incoming solar radiation.

224 7. 3rd place: Zhang

225 The third place approach also used GBRTs, with the differences coming in the data pre-
226 processing for training. Before training, the 11 forecast members were averaged to minimize
227 the training data for efficiency purposes. Zhang trained 2 GBRTs, each on slightly different
228 data. The first was trained on the data from the GEFS model point closest to the prediction
229 point and the second was trained on a weighted average of the nearest 4 GEFS points. The
230 data at each model point p was distance-weighted by the longitude (ϕ) and latitude (λ)
231 distance to each Mesonet site (s) according to Eq 3.

$$w_p = \frac{1}{(0.1 + \sqrt{(\phi_p - \phi_s)^2 + (\lambda_p - \lambda_s)^2})} \quad (3)$$

232 Both models were trained on all 75 of the available features. Zhang also added features for
233 the day of the year and the longitude and latitude and introduced a new feature called “daily
234 differences in downward shortwave solar radiation (ΔS_d).” This feature was defined in Eq.

235 4 as a weighted sum of the downward shortwave solar radiation for each available hour (S_h):

$$\Delta S_d = -0.5S_h(12) - 0.1S_h(15) + S_h(18) + S_h(21) + 0.8S_h(24) \quad (4)$$

236 The final prediction was a weighted vote of the two GBRTs. The weights were determined
237 using cross-validation. Denoting the GBRT trained on the nearest data points as $GBRT_n$
238 and the one trained on the weighted average as $GBRT_{wa}$, the final prediction for a Mesonet
239 site s was

$$final(p) = \frac{0.5 * GBRT_n(s) + GBRT_{wa}(s)}{1.5}. \quad (5)$$

240 8. Student Winner: Louppe and Prettenhofer

241 Our approach is similar in principle to the first place winner (Eustaquio and Titericz) but
242 makes use of robust regression techniques to take uncertainty into account. Our approach
243 comprises two steps: First, we use a non-linear interpolation technique, Gaussian Process
244 regression (also known as Kriging in Geostatistics), to interpolate the coarse GEFS grid to
245 the location of the solar energy production facilities. Second, we use GBRT to predict the
246 daily solar energy output based on the interpolated model and additional spatio-temporal
247 features.

248 Forecast variables measured at the GEFS locations are interpolated non-linearly onto the
249 Mesonet stations using Gaussian Processes (Rasmussen and Williams 2005). More specif-
250 ically, for a given day and a given time period, a Gaussian Process models the value of a
251 given forecast variable (e.g., temperature, humidity, etc) with respect to the location of a
252 station. Uncertainty in the forecast variables is taken into account by modeling the average
253 value over the 11 members of the ensemble, where uncertainty in the ensemble measurements
254 is specified as confidence intervals through the nugget parameter of the Gaussian Process.
255 Using this technique, 75 forecast variables were interpolated per day in the dataset.

256 To enhance our final model, spatio-temporal variables were engineered and added to the
257 75 variables, including:

- Solar features (delta between sunrise and sunset)
- Temporal features (day of year, month of year)
- Spatial features (latitude, longitude, elevation)
- Non-linear combinations of measurement estimates
- Daily mean estimates
- Variance of the measurement estimates, as produced by the Gaussian Processes

The best accuracy was achieved with GBRT. We used the least absolute deviation loss function for robustness and optimized all hyper-parameters on an internal learning set. To further decrease variance of the model, several GBRT instances were built (using different random seeds) and their predictions averaged to form the final predictions. In our opinion, the ability of GBRT to handle outliers in the outputs by using robust loss functions is crucial in this context, due to the volatile nature of solar energy output. Our pipeline was built on top of the Scikit-Learn machine learning library (Pedregosa et al. 2011), offering efficient implementations for both Gaussian Processes and GBRT.

We evaluated the approach on a dataset of daily solar energy measurements from 98 stations in Oklahoma. The results show a relative improvement of 17.17% and 46.19% over the baselines, Spline Interpolation and Gaussian Mixture Models, respectively.

9. Error Analysis

The top contestant methods exhibited similar monthly error characteristics. The monthly mean absolute error for all stations (Fig. 4) follows the average magnitude of solar energy by month with the smallest error in December and January then increasing to the highest error in May and June. All of the contestants have very similar monthly errors with Eustaquio and Titericz consistently having the lowest error. The monthly mean error shows a very

small amount of bias relative to the magnitude of the mean absolute error. Each contestant follows a similar monthly trend in the mean error. Eustaquio and Titericz have a consistently higher mean error than the other models, which is due to the multiplicative factor applied to their results.

Analysis of the station error shows the effects of geography on the predictions. For all contestants, eastern Oklahoma featured generally higher mean absolute errors compared to western Oklahoma with the Oklahoma Panhandle featuring some of the lowest errors (Fig. 5). This solar error distribution mirrors the annual precipitation distribution in Oklahoma. Since the presence of clouds and rain has a large impact on solar energy amounts, and since precipitation location and duration are challenging to predict, this factor is likely a large component of the increased error in eastern Oklahoma. A subset of the stations buck the geographical trend, and analysis of the contest observations shows that some of these stations recorded extended periods of missing data that were filled with the mean solar radiation value for that site. Only a few stations had these discrepancies, so it did not have a significant impact on the overall contest results.

A bootstrap statistical analysis of the forecast errors was performed on the top 4 contestants to determine if there were statistically significant differences in their forecasts. The confidence intervals (Table 2) indicated large amounts of overlap and no statistically significant differences in the top 4 contestants. The scores of the top 7 contestants fall within the confidence interval of the first place winner, and the top 16 contestants fall within the confidence interval of the 4th place winner.

10. Discussion and Lessons Learned

By hosting the forecasting challenge on Kaggle, we dramatically increased the participation and the diversity of the participants from prior years. This diversity includes a significant increase in the international participation as well as participation from people

outside of meteorology. This broader participation was valuable in highlighting meteorological applications for machine learning and data mining. However, it also provided some challenges from the perspective of running a contest with a final session at an AMS annual meeting. Because the winners were largely international participants, they were not able to travel to AMS. Although most of the winners were able to send a pre-recorded video of their talks and there was an informative discussion in the AMS session, future contests could benefit from better use of video technology to engage the winners in discussions in real-time.

The data, evaluation system, and results from the contest have broader applicability for meteorologists in the renewable energy forecasting sector. The contest results showcased a machine learning method, Gradient Boosted Regression Trees (GBRT), that has not been used extensively in the atmospheric science community at this point. Optimized GBRTs have been shown to provide superior performance on this dataset compared to random forests, linear regressions, and neural networks, which were all used by other contestants. In addition to desirable performance characteristics, GBRTs use different optimization functions depending on the problem, and can be tuned for both computational and accuracy constraints. Due to its decision tree roots, GBRT can also be used to extract information about its input variables through variable influence rankings and partial dependence plots. We hope that the results of this contest and the availability of GBRT in both Python and R open-source machine learning libraries encourage the atmospheric science community to apply the algorithm to their existing datasets.

In the spirit of open data and reproducibility, the contest website¹, data, and evaluation system will continue to be available to anyone wishing to compare their approaches against the contest winners. While new submissions will not appear on the leaderboard, people are still invited to compare their algorithm and discuss new findings on the contest forum.

¹<http://www.kaggle.com/c/ams-2014-solar-energy-prediction-contest>

Acknowledgments.

The contest was sponsored by the American Meteorological Society Committees on Artificial Intelligence Applications to Environmental Science, Probability and Statistics, and Earth and Energy, and EarthRisk Technologies. Will Cukierski from Kaggle helped set up the contest website and provided extensive technical support.

REFERENCES

- Breiman, L., 2001: Random Forests. *Machine learning*, **45** (1), 5–32.
- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone, 1984: *Classification and regression trees*. Chapman and Hall/CRC.
- Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. L. Johnson, and M. D. Eilts, 1995: The oklahoma the Oklahoma Mesonet: A technical overview. *Journal of Atmospheric and Oceanic Technology*, **12**, 5–19.
- Freund, Y. and R. E. Schapire, 1995: A desicion-theoretic generalization of on-line learning and an application to boosting. *Computational learning theory*, Springer, 23–37.
- Friedman, J. H., 2001: Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, **29** (5).
- Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Y. Zhu, and W. Lapenta, 2013: Noaa’s second-generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc.*, **94**, 1553–1565.
- Lakshmanan, V., K. L. Elmore, and M. B. Richman, 2010: Reaching scientific consensus through a competition. *Bull. Amer. Meteor. Soc.*, **91**, 1423–1427.

- Lakshmanan, V. and J. S. Kain, 2010: A gaussian mixture model approach to forecast verification. *Wea. Forecasting*, **25**, 908–920.
- McPherson, R. A., et al., 2007: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. *Journal of Atmospheric and Oceanic Technology*, **24**, 301–321.
- Nelder, J. A. and R. Mead, 1965: A simplex algorithm for function minimization. *Computer Journal*, **7**, 308–313.
- Pedregosa, F., et al., 2011: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.
- Rasmussen, C. E. and C. K. I. Williams, 2005: *Gaussian Processes for Machine Learning*. The MIT Press.
- Saha, S. and Coauthors, 2010: The NCEP climate forecast system reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057.
- Shafer, M. A., C. A. Fiebrich, D. S. Arndt, S. E. Fredrickson, and T. W. Hughes, 2000: Quality assurance procedures in the quality assurance procedures in the Oklahoma Mesonet-work. *Journal of Atmospheric and Oceanic Technology*, **17**, 474–494.
- Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operation forecast system. *Tellus A*, **60**, 62–79.

371 **List of Tables**

| | | | |
|-----|---|--|----|
| 372 | 1 | Variable influence rankings for the 2nd place gradient boosting algorithm. | 17 |
| 373 | 2 | 95% bootstrap confidence intervals for each of the top 4 contestants. | 18 |

TABLE 1. Variable influence rankings for the 2nd place gradient boosting algorithm.

| Name | Percent |
|----------------|----------------|
| dswrf (21 UTC) | 20.9% |
| dswrf (18 UTC) | 13.1% |
| dswrf (00 UTC) | 11.5% |
| dswrf (15 UTC) | 4.2 % |
| pwat (21 UTC) | 3.8 % |
| pwat (15 UTC) | 3.7 % |
| pwat (18 UTC) | 3.6 % |
| Month | 3.5 % |
| pwat (00 UTC) | 3 % |
| pwat (12 UTC) | 2 % |

TABLE 2. 95% bootstrap confidence intervals for each of the top 4 contestants.

| Contestant | 95% Confidence Interval (MJ m ⁻²) |
|----------------------------|--|
| 1. Eustaquio and Titericz | (2.028, 2.180) |
| 2. Lazorthes | (2.044, 2.211) |
| 3. Zhang | (2.077, 2.224) |
| 4. Louppe and Prettenhofer | (2.082, 2.244) |

374 List of Figures

| | | | |
|-----|---|---|----|
| 375 | 1 | Map of the grid points from the GEFS (blue) and the Oklahoma Mesonet | |
| 376 | | station sites (red) superimposed on a NASA Blue Marble satellite image back- | |
| 377 | | ground. | 20 |
| 378 | 2 | Data pre-processing to handle the Mesonet stations being on a different grid | |
| 379 | | than the GEFS model. | 21 |
| 380 | 3 | Partial dependence plot derived from a gradient boosting model. | 22 |
| 381 | 4 | Monthly mean absolute error and mean error for each of the top 4 contestants. | 23 |
| 382 | 5 | Mean absolute error at each Mesonet site for the top 4 contestants. | 24 |

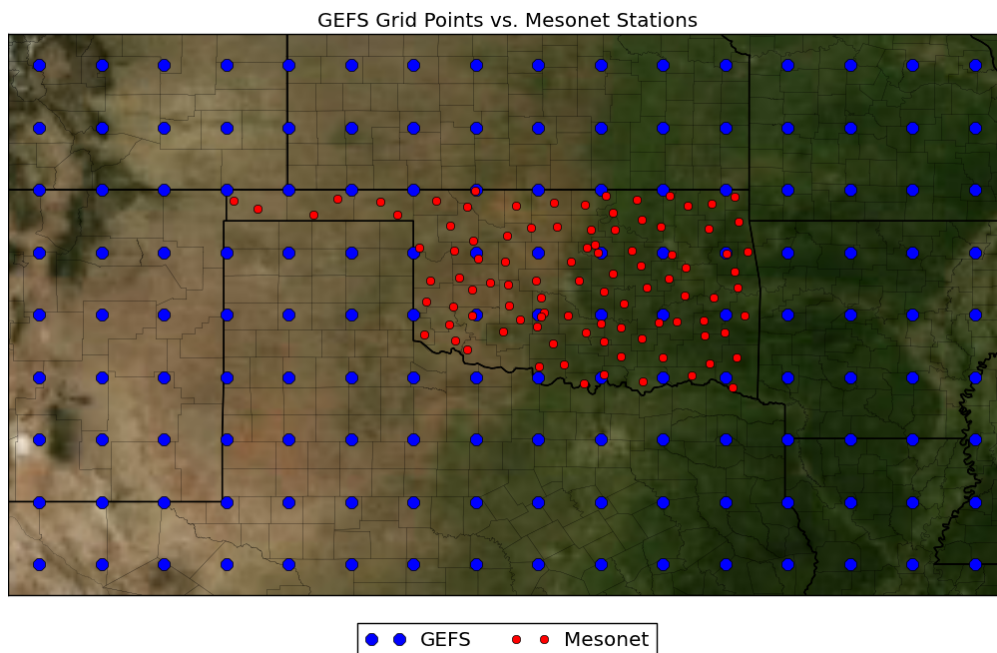


FIG. 1. Map of the grid points from the GEFS (blue) and the Oklahoma Mesonet station sites (red) superimposed on a NASA Blue Marble satellite image background.

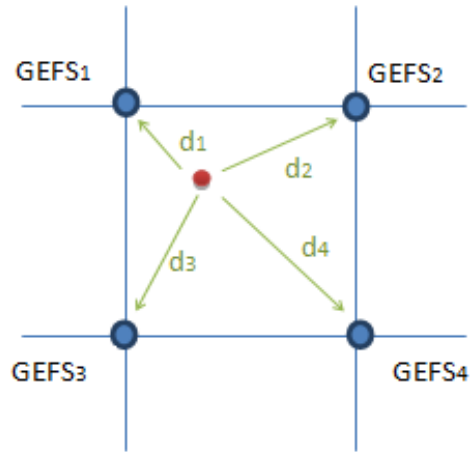


FIG. 2. Data pre-processing to handle the Mesonet stations being on a different grid than the GEFS model.

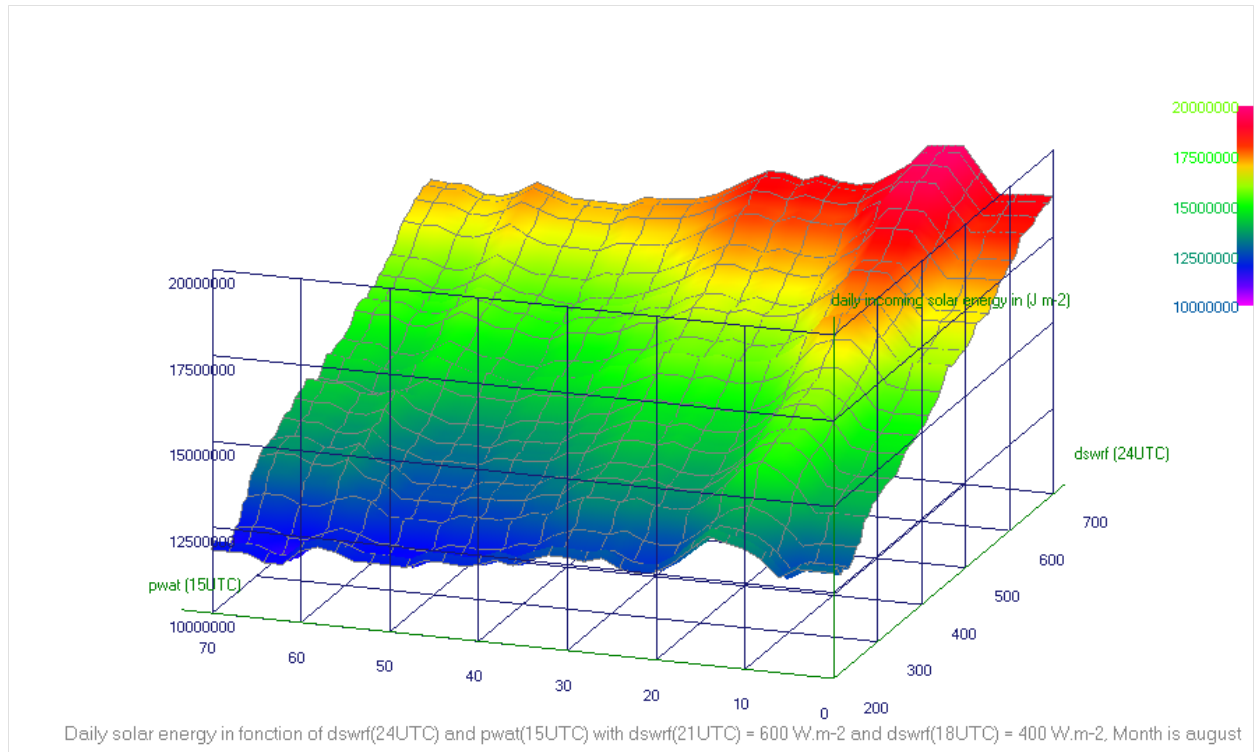


FIG. 3. Partial dependence plot derived from a gradient boosting model.

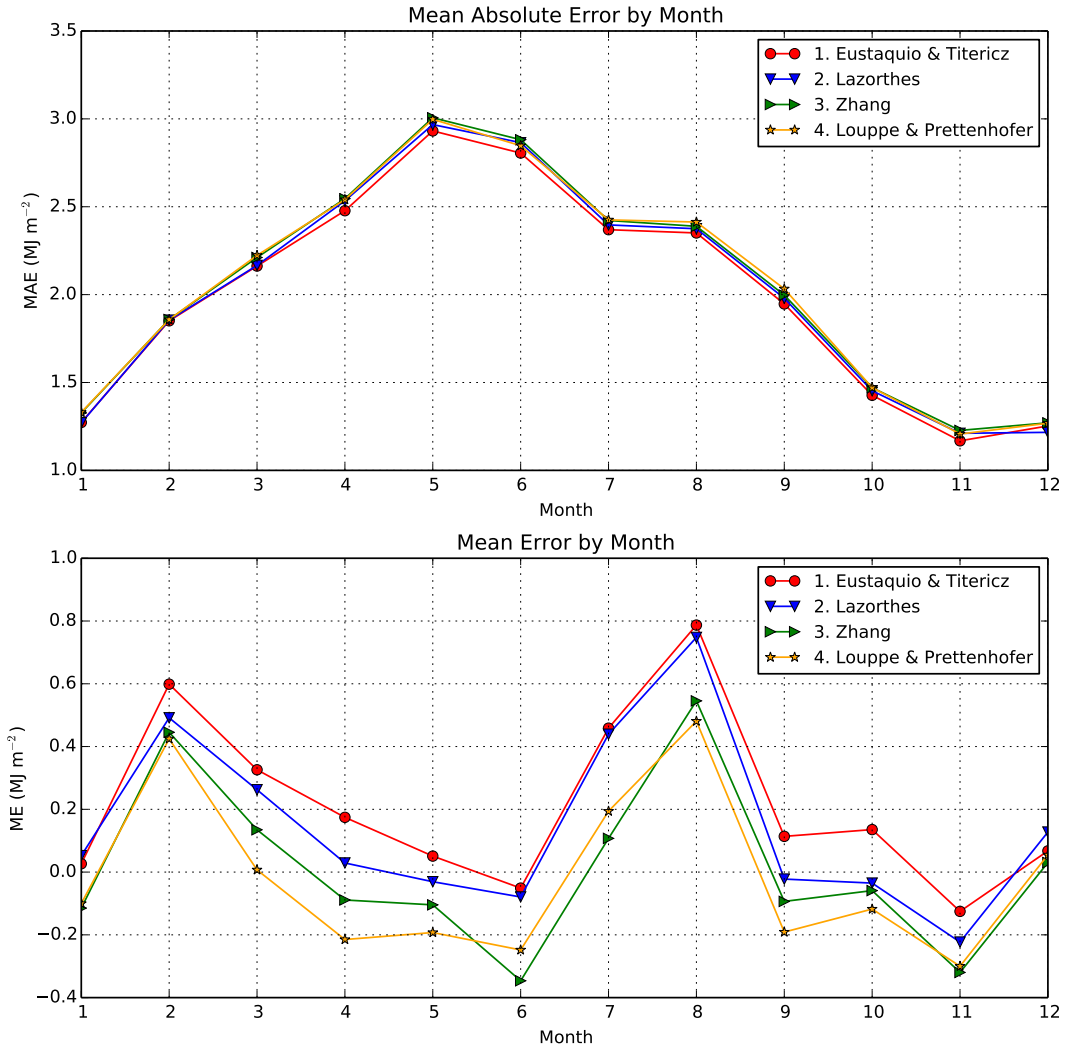


FIG. 4. Monthly mean absolute error and mean error for each of the top 4 contestants.

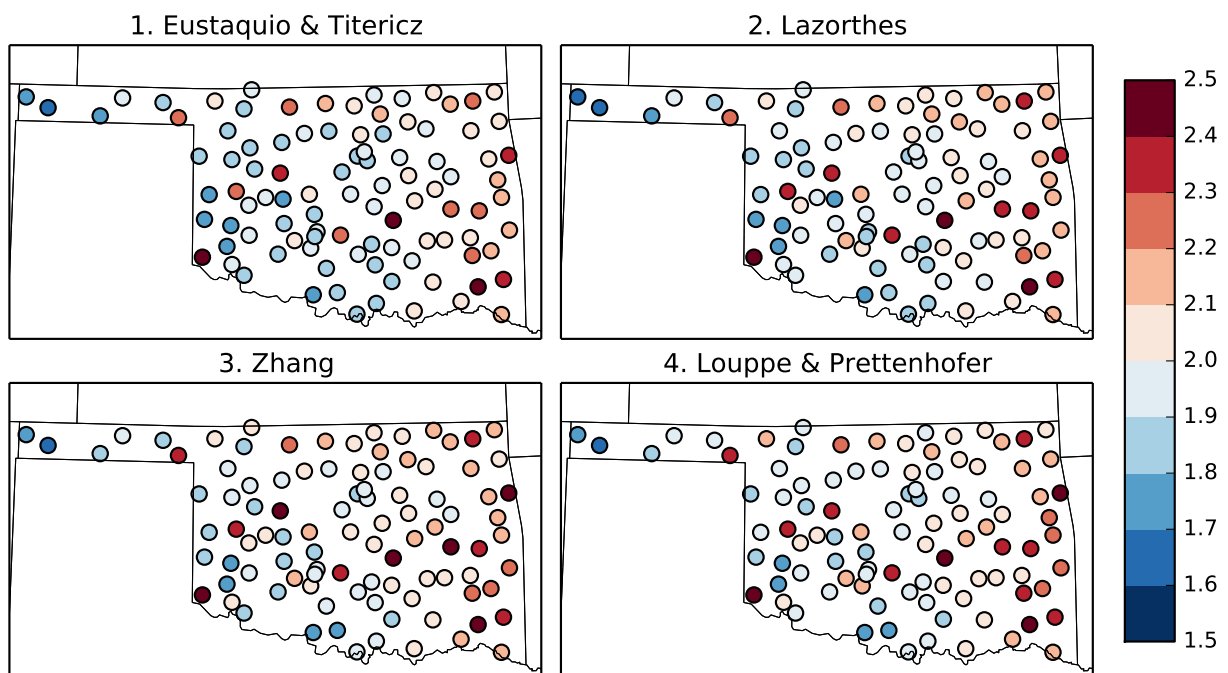


FIG. 5. Mean absolute error at each Mesonet site for the top 4 contestants.