Prediction of Lifted Condensation Level and Solar Irradiance using Machine Learning

Anmol Dash

Adlai E. Stevenson High School Lincolnshire, Illinois

anmol@live.com

Abstract—This paper explores the creation and use of machine learning models to predict solar irradiance and lifted condensation level(LCL) from local weather data and thermal images. The datasets were taken from the NCEI Local Climatological database, a pyrometer setup at Argonne National Laboratory, and a set of thermal images, also taken at Argonne National Laboratory. Several models from Pytorch, Tensorflow,/Keras, and Scikit-learn were created in order to predict the LCL pressure and LCL temperature, the two values that could be calculated, as well as the net irradiance(the amount of radiance that reaches the ground) and global irradiance(the total amount of irradiance). For LCL pressure, the best performing model (accuracy=.997)was a Keras Functional(accuracy = .977) used the relative humidity, the sea level pressure, and the wet bulb temperature along with the thermal images. For LCL temperature, the best performing model was a Keras Sequential model that only used relative humidity, sea level pressure, and wet bulb temperature. For net irradiance, the best performing model(accuracy=.948)used the data collected along with the pyrometer along with the thermal images, and the best performing model for the global irradiance(accuracy=.925) took all the data. The models require very little data in order to calculate values that normally take specialized equipment. In the future, these could potentially be integrated into a system in their respecive areas of need in order to faster and more accurately compute needed meteorological information.

Index Terms—Lifted Condensation Level, Convective Condensation Level, Solar Irradiance, Net Irradiance, Global Irradiance

I. INTRODUCTION

This paper explores the potential for the prediction of lifted condensation level and solar irradiance through Machine Learning(ML) techniques. Lifted condensation level, or the height at which an air parcel reaches 100% humidity, is vital for the prediction and tracking of storms[1]. Lifted condensation is often simply known as the base of a cloud. Rather than being calculated, lifted condensation has often been used in other machine learning models, including one for global horizontal irradiance[2]. The other focus of this paper, solar irradiance, is extremely important for the efficient operation of solar power plants, and to predict/and or forecast it is crucial to setting up new plants[3]. With climate change becoming an ever bigger issue, and with solar power plants providing a cheap and reliable alternative, the calculation of solar irradiance become ever more important. With the advancement of machine learning technology, and potential for rapid calculation of these measurements without the need for specialized equipment, the potential utilization of both solar irradiance and lifted condensation level is great. While this is a novel approach, similar studies have been conducted on convection to predict storms over the Indian Ocean, using several different meteorological measurements, like humidity, wind speed, and pressure, and another used nowcasted Solar Irradiance within 15 minute to 1 hour intervals, using deep learning techniques to process several different images, as well as ground meteorological measurements[4, 5].Similar to this paper, the previous papers used Scikit-learn modules and deep learning techniques to create different models. In particular, solar irradiance forecasting is a well-researched topic, and has had many potential avenues explored[5, 6, 7]. Most of these strategies heavily rely on images, both thermal and otherwise, in order to calculate the Solar Irradiance. In this paper, rather than images, the arrays derived from the measurements the thermal cameras use to create the thermal images will be used to calculate the solar irradiance and lifted condensational level. The three main libraries for the execution of this are Tensorflow/Keras, Pytorch, and Scikit-learn. As mentioned before, Scikit-learn has been used in similar studies to predict similar values. Tensorflow/Keras has also been used to model climate and convection in other studies, and Pytorch has been used for image processing in ocean climate models[8, 9]. This paper is structured in four parts. First, the datasets, where we will explore the data used to train and test the models. Second, the models, where we will analyze the different models used to try to predict both irradiance and lifted condensation level. Third, the results, where we will review the results of the models. Finally, the discussion and conclusion, where we will review the results and see how they compare with the expectations laid out in the beginning of the paper.

II. DATASETS

A. Local Climatological Data

The local meteorological data was taken from the National Centers for Environmental Information Climate Data Online database for Local Climatological Data. The data was taken from 4/15/2021 to 4/23/2021 at O'Hare Airport(WBAN:94846) .The measurements taken were Hourly Dew Point temperature(F), Hourly Dry Bulb temperature(F), Hourly Relative Humidity,Hourly Sea Level pressure(inHg), Hourly Visibility(km),Hourly Wet Bulb temperature(F), Hourly Wind Direction(deg), Hourly Wind Speed(dm/s) The dataset had 310 entries.

B. Thermal Images

Taken using a MOBOTIX M16 AllAround Dual™ thermal camera at Argonne National Laboratory from 4/15/2021 to 4/23/2021. The dataset contains 1524 thermal images, as well as corresponding thermal measurements which were formatted as CSV arrays. These arrays had a variety of sizes, but for the purpose of most models, they were reshaped into a 252x336 NumPy 2D array.

C. Pyrometer

Taken using a pyrometer at Argonne National Laboratory from 4/15/2021 to 4/23/2021, the dataset contained Net Solar Irradiance(W/m) and Global Solar Irradiance(W/m) , as well as other measurements taken at the site, which were temperature(C), wind speed(cm/s) , wind direction(degrees), relative humidity(%), dew point(C), soil temperature(C), pressure(kPa), and precipitation(mm)

D. Combining the Datasets

In order to create a training dataset out of the original data for LCL pressure, LCL temperature, Net Solar Irradiance, and Global Solar Irradiance for the different models, each thermal image was aligned with the dataset of the entries taken from the other datasets. Since both the Local Climatological Data and the additional data from the pyrometer setup were taken on one hour intervals, the thermal images were aligned with the image taken in the same hour, duplicating the rows to ensure that all the thermal images have their own row. Unfortunately, the Local Climatological Data also had many rows with NaN values, and they were decreased through pandas dropna() function. The entries dropped differed based on the model that was created, and so the base dataset was still at 1524 entries.

E. Prediction Adjustment

- LCL Pressure and Temperature
  - Using MetPy's function for calculating LCL pressure and LCL temperature, the Local Climatological Data was used to calculate the LCL pressure and temperature using the Hourly Dry Bulb temperature, Hourly Dew Point temperature, and Hourly Station pressure. All of these were removed from the training and testing dataset of all the models.
- Solar Irradiance
  - When training and testing a model to predict net global irradiance, both were removed training testing dataset.

III. MODELS

A. Scikit-learn regression

The first types of model used were six different Scikit-regression algorithms.

- Support Vector Regression(SVR)
- Stochastic Gradient Descent Regressor(SGDRegressor)
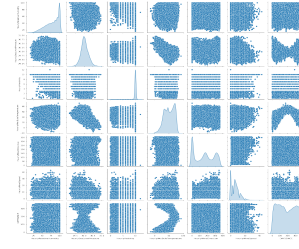- Bayesian Ridge
- LassoLars

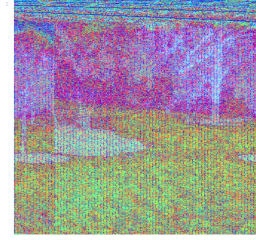Fig. 1. Seaborn Graph of Local Climatological Dataset

Fig. 2. Image created from temperature array. Displays the location where they were taken, as well as the contrast between the sky, grass, and trees.

- Automatic Relevance Determination Regression(ARD)
- Passive Aggresive Regressor(PAR)

All were solely used in the prediction of LCL temperature. Scikit-learnregression do not accept multiple inputs, so use of the thermal images was not possible.

B. PyTorch

The Pytorch model was solely used for the prediction of net solar irradiance. The model tested was a Deep Neural Network(DNN), with 5 hidden layers after concatenation of the two inputs, which were the data collected by the equipment near the pyrometer and the thermal images. The thermal images had 2 hidden layers of processing before concatenation with the data collected by the equipment. All the layers, with
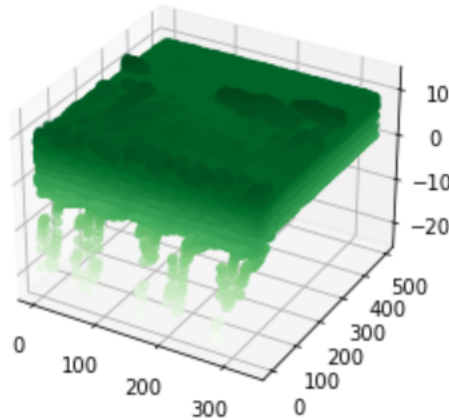

Fig. 3. Graph created from temperature array. Once again the contrast is displayed, with the y-values marking temperature, and the x and z coordinates marking where each dot would have been on the array

the exception of the output, had a ReLu activation. The loss was calculated using the mean squared error(MSELoss), and the optimizer was Stochastic Gradient Descent. The learning rate was .0001 and the model was always trained with 1000 epochs.

C. Tensorflow/Keras

1) Keras Sequential: The Keras Sequential models were used for calculating the LCL pressure and temperature, as well as the Net Irradiance and Global Irradiance. For all three, the model architecture consisted of three hidden layers with ReLu activation. The model's optimizer was Adam, the learning rate was .001 and the loss was mean absolute error. The validation split was .2, and the epochs used to train the model were 100 for everything except the global solar irradiance.

2) Keras Functional: The Keras Functional models were the most complex of all the models. It takes two inputs when calculating both the two LCL values and the net solar irradiance, one for the associate dataset, and the other for the 252x336 temperature NumPy array. The model architecture consisted of two input layers for the two inputs, and three hidden layers before concatenation. After concatenation, the values go through 5 hidden layers, with all layers, including the ones before concatenation, having ReLu activation. The last layer, as in the Sequential model, contained no activation. The loss for the LCL values was mean squared error, and the optimizer was RMSprop. The metric for the LCL values was accuracy, which, despite being a classification metric, was the metric that seemed to work the best after multiple trials. For the net irradiance, both the loss and the metric was mean absolute error, and the optimizer was RMSprop. Both of them had a minimum number of training epochs of 1000.

IV. RESULTS

A. LCL temperature

1) Scikit-learn: The six Scikit-learn models all performed relatively badly on the training set. All the models used all the features provided in the Local Climatological Dataset.

- SVR- accuracy: 0.196, MSE: 1.675
- SGDRegressor- accuracy: -4.07694e+34, MSE: 1.021
- Bayesian Ridge- accuracy: 0.34, MSE: 1.52
- LassoLars - accuracy: 1.11e-16, MSE: 1.87
- ARDRegression - accuracy: 0.33, MSE: 1.53
- PassiveAggressiveRegressor - ACC: -13.68, MSE= 7.67

2) Keras Sequential: The Keras Sequential model was successful in predicting the LCL temperature from only the Hourly Sea Level pressure, Hourly Relative Humidity, and Hourly Wet Bulb temperature values from the Local Climatological Dataset with an accuracy of .98 and a MSE of .304.

B. LCL pressure

1) Keras Sequential: The Keras Sequential model, with only the Hourly Sea Level pressure, Hourly Relative Humidity, and Hourly Wet Bulb temperature values. obtained an accuracy of .755 and a MSE of .8633

2) Keras Functional: The Keras Functional model, with both the 252x336 model and the Hourly Sea Level pressure, Hourly Relative Humidity, and Hourly Wet Bulb temperature values, obtained an accuracy of .977 and a MSE of .576

C. Net Solar Irradiance

1) Keras Functional: The model obtained an accuracy of .94 and a MSE of 38.6 with the dataset obtained alongside the pyrometer..

2) Keras Sequential: The Keras Sequential model, with the Local Climatological dataset combined with the net irradiance . obtained an accuracy of .997 and a MSE of 9.77

3) Pytorch: The Pytorch models, with the entire pyrometer dataset and thermal arrays, obtained an accuracy of -.136 and a MSE of 254.69

D. Global Solar Irradiance

1) Keras Sequential-Local Climatological Data: The Keras Sequential model, using the Local Climatological Data, obtained a final result of accuracy: .95, MSE: 54.167.

2) Keras Functional: The Keras Functional model obtained a final accuracy of .9019 and a MSE of 60.53.

3) Keras Sequential-Pyrometer: The Keras Sequential model using the pyrometer dataset obtained an accuracy of .925 and an MSE of 83.055
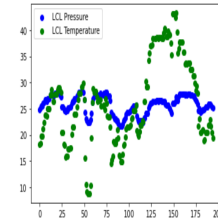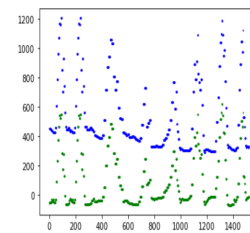

Fig. 4. x= time, y= either LCL pressure or temperature

Fig. 5. x= time. The green is global irradiance and the blue is net irradiance

V. DISCUSSION

A. Lifted Condesation Level Predictions

1) LCL Temperature: LCL temperature was the first measurement successfully predicted by any model over the course of the entire summer. As the result, the model for it is far less complex than the ones made later, like the Keras Functional model created for net solar irradiance. However, the Keras Sequential model is still highly accurate with the
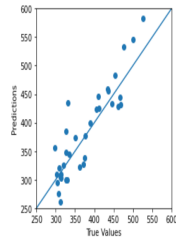
Fig. 6. True Values vs. Predictions for the global Irradiance Keras Sequential model

correct inputs, and so has provided a framework for the other models. The three inputs needed (Hourly Sea Level pressure, Hourly Relative Humidity, and Hourly Wet Bulb temperature) were not selected because of a negative impact on prediction accuracy with the addition of other features. Rather, the baseline accuracy was still .977, even before the removal of the other features. The minimum number needed are the three inputs mentioned above. The failure of the Scikit-learn model can easily be attributed to the fact that the pattern for LCL temperature, as shown in the diagram above, is not linear, but rather repeats itself throughout the entire dataset. In addition, because there was a spread of values in the LCL temperature dataset, as opposed to the LCL pressure, the use of the mean absolute error as the loss function was probably the best choice, lending once more the relatively high final accuracy.

While this model may be highly accurate with the data given, it also presents significant challenges with regards to actual implementation. The first is the features themselves. While relative humidity is a relatively novel feature for predicting LCL temperature, but Sea Level pressure and Wet Bulb temperature are very closely related to Station pressure and Dry Bulb temperature. In the absence of equipment to measure either one, though, this model is a highly accurate method for correctly obtaining the LCL temperature. Secondly, the use of the Sea Level pressure also creates the issue of accuracy farther away from sea level. With no equipment to test that issue though, there is no practicable way to resolve it. Perhaps further study, or possibly the testing of higher altitude data, would find the answer.

2) LCL Pressure: LCL pressure was the second model created that obtained a high accuracy. As with LCL temperature, the features necessary to predict the LCL pressure were not selected because of the negative impact of the other features, but rather the accuracy of the model created staying the same, with the exact same parameters, as each feature was removed. The failure of the Sequential model is interesting, as well as the complexity of the Functional model found to be most successful. To the first point, the LCL pressure, as shown above, is far more repetitive than the LCL temperature, but also contained within a far smaller range. I hypothesize that the reason the Sequential model failed was simply because it could not predict exactly within such a small range. To the same end, the complexity of the Functional model perhaps

creates the potential for more precise outputs, contributing to the higher accuracy measurement. The use of the accuracy metric, which is a classification metric, rather than a regression metric, was an initial error that went unnoticed. However, when changed, the model did not function as it did before, so it remained. Perhaps further study can explore the importance of the accuracy metric, or why no other metric functions as well.

3) Observations for Both: When a Functional model was created to predict LCL temperature with the temperature array, the accuracy dropped sharply as compared to the Sequential model. In contrast, when the second input was removed, but the model architecture stayed the same, the accuracy was comparable to the Sequential model. With LCL pressure, the removal of the temperature array created a sharp drop in accuracy. This almost opposing reaction to the same input is curious, especially since the two roughly align in distribution over time. Overall, the models worked fast and well, and are an example of what could be achieved with proper utilization of machine learning techniques

B. Solar Irradiance Predictions

1) Net Irradiance: Of the two irradiance measurements, Net Irradiance has more potential applications, but also far more success with regards to prediction, at least as compared to Global Irradiance. The first model to consider is the Pytorch model. It's failure can easily be attributed to a few factors. First, the SGD Regressor, since it maintains a constant learning rate, is perhaps not suited to a dataset that is so non-linear. In addition, the number of hidden layers and the concatenation is perhaps too complex, resulting in a model that is fundamentally non-functional. Also, the use of mean squared error as a loss functions could have possibly affected the final accuracy of the model to predict a dataset that has a relatively high variability compared to other datasets tested before, like the LCL pressure. The success of the Keras sequential can also be called into question, as the ASOS data and the net irradiance measurements don't precisely line up, and most of the irradiance values were repeated several times throughout the dataset. Combined with a relatively complex model, and the possibility of an erroneous model that only performs well on the testing dataset, but will not accurately predict irradiance outside of the dataset. Finally, the Keras Functional model's success is surprising, but also perhaps the most reliable. The Keras Functional model, unlike the other models, has also been tested on several different testing datasets and maintained the same accuracy. The only problem with this is the dataset. In order to predict the irradiance, 3 different measurements need to be taken for sigma direction and wind direction, and also 2 different measurements for air tempearature. It is unclear how exactly the three measurements were taken, if they were taken on different devices, or are simple measurements over the course of an hour from the same device. The best assumption, and most likely the most accurate one, is that the multiple measurements came over the course of an hour- the hour over which there is a single irradiance value.

2) Global Irradiance: Global Irradiance is both the larger and more complex one of the two types of Irradiance. With a far larger distribution of values, the global irradiance also obtained a far lower accuracy than the net irradiance. The Keras Sequential Model's success still faces the same issues seen in its success in predicting net irradiance. The Keras Functional model, having an almost identical architecture to the net irradiance model, obtains, on average, a far lower accuracy. It is also far less reliable, obtaining a wide range of accuracies across multiple testing sets.

3) Observations for Both: Overall, net irradiance seems to be the more reliable result of prediction, though the Keras Sequential models and the Keras Functional model for predicting global irradiance both have acceptable levels of accuracy. The successful use of the Sequential model and the necessary absence of the thermal images once suggests that the thermal image may be more important in identifying particular features beyond the scope of this paper, or perhaps that the variation found in the thermal images is unfavorable to the prediction of the global irradiance.

VI. Conclusion

In summary, through machine learning, models to accurately calculate and predict the LCL pressure and temperature, as well as the global and net irradiance within reasonable limits have been created. The models needed were varied, with LCL pressure and net irradiance requiring highly complex Keras Functional models, while LCL temperature and global irradiance simply required Keras Sequential models. In addition, while the Local Climatological dataset could be reduced to just three features(relative humidity, sea level pressure, and wet bulb temperature), the pyrometer dataset needed multiple measurements of each feature. The origin of these multiple measurements is unknown, because these readings were taken a while before I started, but they are likely taken at multiple times throughout the hour that the irradiance was measured. Throughout the paper, the models that best suited the dataset, not necessarily the most trained or most complex model architectures succeeded. This emphasizes the importance of exploring different models in order to develop the higher accuracy. Some models, and by association, rejected the use of thermal images, suggesting a link between the thermal images and, in particular, LCL pressure and net irradiance that warrants further exploration. In addition, the failures of Pytorch and Scikit-learn need to be investigated further. While the low accuracy could be due to human error, there also could be a significant issue with the dataset. Once again, future exploration is warranted.

References

[1] Brown M, Nowotarski CJ. The Influence of Lifting Condensation Level on Low-Level Outflow and Rotation in Simulated Supercell Thunderstorms. J Atmos Sci. 2019;76(5):1349-1372. doi:10.1175/JAS-D-18-0216.1.

[2] Ding Y, Xu C, Fang C, Xiang Z, Hai Z. Very-short term forecast of global horizontal irradiance based on ground-based sky imager and lifted condensation level calculation. China Int Conf Electr Distrib CICED. 2014;2014-December 973-978. doi:10.1109/CICED.2014.6991849

[3] Wittmann M, Breitkreuz H, Schroedter-Homscheidt M, Eck M. Case studies on the use of solar irradiance forecast for optimized operation strategies of solar thermal power plants. IEEE J Sel Top Appl Earth Obs Remote Sens. 2008;1(1):18-27. doi:10.1109/JSTARS.2008.2001152

[4] Nueve E, Jackson R, Sankaran R, Ferrier N, Collis S. WeatherNet: Nowcasting Net Radiation at the Edge.

[5] Ukkonen P, Mäkelä A. Evaluation of Machine Learning Classifiers for Predicting Deep Convection. J Adv Model Earth Syst. 2019;11(6):1784-1802. doi:10.1029/2018MS001561

[6] Lou S, Li DHW, Lam JC, Chan WWH. Prediction of diffuse solar irradiance using machine learning and multivariable regression. Appl Energy. 2016;181:367-374. doi:10.1016/J.APENERGY.2016.08.0.

[7] Dazhi Y, Jirutitijaroen P, Walsh WM. Hourly solar irradiance time series forecasting using cloud cover index. Sol Energy. 2012;86:3531-3543. doi:10.1016/j.solener.2012.07.029

[8] Gentine P, Pritchard M, Rasp S, Reinaudi G, Yacalis G. Could Machine Learning Break the Convection Parameterization Deadlock? Geophys Res Lett. 2018;45(11):5742-5751. doi:10.1029/2018GL078202

[9] Ishida K, Tsujimoto G, Ercan A, Tu T, Kiyama M, Amagasaki M. Hourly-scale coastal sea level modeling in a changing climate using long short-term memory neural network. Sci Total Environ. 2020;720:137613. doi:10.1016/J.SCITOTENV.2020.137613