

A Deep Machine Learning Algorithm to Optimize the Forecast of Atmospherics

Alexandria M. Russell, Randall J. Alliss, Billy D. Felton

*Northrop Grumman Mission Systems
7555 Colshire Dr
McLean, VA 22102*

Abstract

Space-based applications from imaging to optical communications are significantly impacted by the atmosphere. Specifically, the occurrence of clouds and optical turbulence can determine whether a mission is a success or a failure. In the case of space-based imaging applications, clouds produce atmospheric transmission losses that can make it impossible for an electro-optical platform to image its target. Hence, accurate predictions of negative atmospheric effects are a high priority in order to facilitate the efficient scheduling of resources.

This study seeks to revolutionize our understanding of and our ability to predict such atmospheric events through the mining of data from a high-resolution Numerical Weather Prediction (NWP) model. Specifically, output from the Weather Research and Forecasting (WRF) model is mined using a Random Forest (RF) ensemble classification and regression approach in order to improve the prediction of low cloud cover over the Haleakala summit of the Hawaiian island of Maui. RF techniques have a number of advantages including the ability to capture non-linear associations between the predictors (in this case physical variables from WRF such as temperature, relative humidity, wind speed and pressure) and the predictand (clouds), which becomes critical when dealing with the complex non-linear occurrence of clouds. In addition, RF techniques are capable of representing complex spatial-temporal dynamics to some extent.

Input predictors to the WRF-based RF model are strategically selected based on expert knowledge and a series of sensitivity tests. Ultimately, three types of WRF predictors are chosen: local surface predictors, regional 3D moisture predictors and regional inversion predictors. A suite of RF experiments is performed using these predictors in order to evaluate the performance of the hybrid RF-WRF technique. The RF model is trained and tuned on approximately half of the input dataset and evaluated on the other half. The RF approach is validated using in-situ observations of clouds.

All of the hybrid RF-WRF experiments demonstrated here significantly outperform the base WRF local low cloud cover forecasts in terms of the probability of detection and the overall bias. In particular, RF experiments that use only regional three-dimensional moisture predictors from the WRF model produce the highest accuracy when compared to RF experiments that use local surface predictors only or regional inversion predictors only. Furthermore, adding multiple types of WRF predictors and additional WRF predictors to the RF algorithm does not necessarily add more value in the resulting forecasts, indicating that it is better to have a small set of meaningful predictors than to have a vast set of indiscriminately-chosen predictors. This work also reveals that the WRF-based RF approach is highly sensitive to the time period over which the algorithm is trained and evaluated. Future work will focus on developing a similar WRF-based RF model for high cloud prediction and expanding the algorithm to two-dimensions horizontally.

1. INTRODUCTION

With ever-increasing amounts of space-based data being generated and transmitted by the general public, scientists, and the military, modern society is increasingly reliant on high-performance satellite communications. As users continue to demand more data, the existing communications infrastructure will have to expand to meet the demands. Radio Frequency signals have been relied on exclusively and successfully to communicate with spacecraft since satellite communications began nearly 60 years ago, but there are limitations that may prevent radio frequency communications from fully meeting future requirements. These technological, regulatory, and financial limitations may be alleviated, in part, by Free-Space Optical Communications (FSOC) systems. There are several advantages to using FSOC to meet future communications requirements. In particular, data can be transmitted through free-space via lasers at very high data rates of multi-Gb/s over long distances. Optical beams are also very narrow making them much less susceptible to interference or interception than radio frequency signals. Additionally, unlike radio frequencies, the optical spectrum is unregulated. Finally, optical communications systems are relatively small and potentially much less expensive than comparable radio frequency systems, particularly for space missions.

The ultimate realization of practical, high-availability FSOC systems, however, will depend on how well they can mitigate the impacts of atmospheric effects, primarily cloud cover and optical turbulence (OT). Clouds are the largest source of atmospheric attenuation for space-to-ground optical communications, often producing transmission losses of several decibels (dBs) to several tens of dBs. Without impractically large link margins, most clouds are generally considered blockages to FSOC links.

Since atmospheric phenomena such as clouds are the major limiting factor to the success of FSOC, ground sites for FSOC systems are preferentially located in areas with little to no atmospheric interference (i.e. dry and largely cloud-free sites at high altitudes). One such location is the Haleakala peak on the Hawaiian island of Maui where a semi-persistent low-level temperature inversion (e.g. Fig. 1) often prevents clouds from encroaching upon the summit. Due to its height, remoteness, and atmospheric characteristics Haleakala was selected for telescopes such as the Mees Solar Observatory, the Advanced Electro-Optical System, and the Air Force Maui Optical Station.

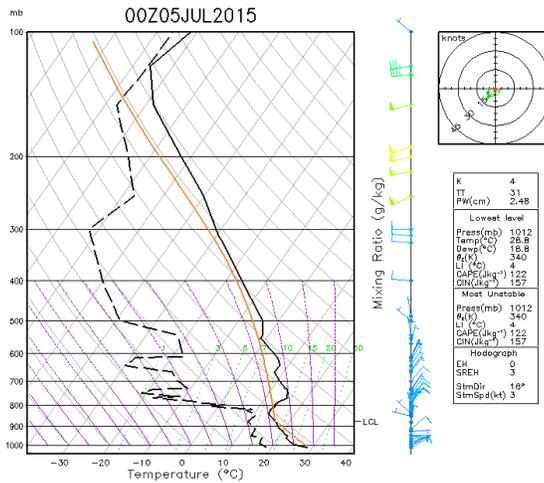


Fig. 1. Hilo Skew-T Log-P diagram for 0000 UTC on July 5, 2015, with temperature (solid black line) and dew point temperature (dashed black line).

The height of the trade-wind inversion around Maui varies between 1500 and 2500 meters above mean sea level, meaning the summit of Haleakala (at about 3055 meters) is usually above the cloud tops. However, the inversion can sometimes be higher and/or weaker, thereby allowing clouds to envelop the summit and interfere with imaging and optical communications operations. In addition, the unique terrain of Haleakala, with the summit sitting at the western end of a long, deep valley (the Haleakala caldera), can at times enhance convergence and force clouds up and over the summit even when the trade wind inversion is somewhat below the summit. Hence, regardless of FSOC ground-site location conditions, accurate forecasts of meteorological phenomena are foundational to the mitigation of atmospheric interference.

Global weather forecasts are generally too coarse in spatial resolution to provide enough accuracy at the highly local scale that is required for FSOC. Therefore, high-resolution weather forecasts on the scale of kilometers to sub-

kilometers, depending on the local meteorology, are often required. State-of-the-art Numerical Weather Prediction (NWP) models offer high accuracy forecasts of atmospheric conditions on the spatial and temporal scales of interest to FSOC. NWP systems use sophisticated dynamical models to predict meteorological variables for points in a regional or global grid. The resulting forecasts can be used to develop decision aids for station hand-off, maintenance-scheduling and other FSOC operational activities.

Even with the advanced capabilities of NWP systems, forecasting atmospheric interference on a local scale at high temporal resolutions remains a challenging task. One way to further improve upon NWP forecasts is to develop hybrid Machine Learning (ML) forecasting techniques based on NWP model output. ML tools such as Artificial Neural Networks and ensemble learning methods are theoretically able to improve suboptimal forecasts by 1) compensating for biases in the training dataset (in this case NWP forecasts) and 2) utilizing aspects of the training dataset that are highly significant, all while avoiding over-fitting the data. In fact, hybrid ML-NWP techniques have been shown to outperform base NWP forecasts in a variety of settings [1,2,3,4,5]. Here, an algorithm based on a ML technique, Random Forest (RF), is developed to predict low cloud cover with a lead-time of up to 48 hours over the Haleakala summit using large amounts of sub-mesoscale NWP output. This approach is validated using local in-situ observations of low cloud cover.

The general outline of this paper is as follows. Section 2 describes the NWP model and validation dataset. Section 3 describes the Machine Learning approach based on NWP model output for low cloud cover prediction. Section 4 provides the results of several RF experiments with different types of NWP variables as well as the performance of the native NWP model output. A summary of the results and a discussion of this work's limitations and future directions are provided in Section 5.

2. MODELS AND DATA

a. WRF Model Configuration

The Weather Research and Forecasting (WRF) is a next-generation fully three-dimensional (3D) physics-based mesoscale NWP model. WRF model version 3.6.1 [6] is used here to produce forecasts of cloud cover over the summit of Haleakala. The WRF model configuration used here consists of an outer domain run at 9-km resolution that contains much of the central Tropical Pacific Ocean, a regional 3-km grid centered on the Hawaiian Islands, an inner 1-km grid that contains the island of Maui and neighboring islands, and a 1/3-km resolution grid centered on the summit of Haleakala (Fig. 2). All domains have 81 vertical levels with a resolution of approximately 50-100 m below 2 km above ground level (AGL), 150-250 m for 2–13 km AGL, and 500 m up to the model top (50 millibars). The high horizontal and vertical resolutions described here allow for more accurate forecasting of the fine-scale atmospheric circulations around Haleakala whose local meteorology is heavily influenced by the complex topography.

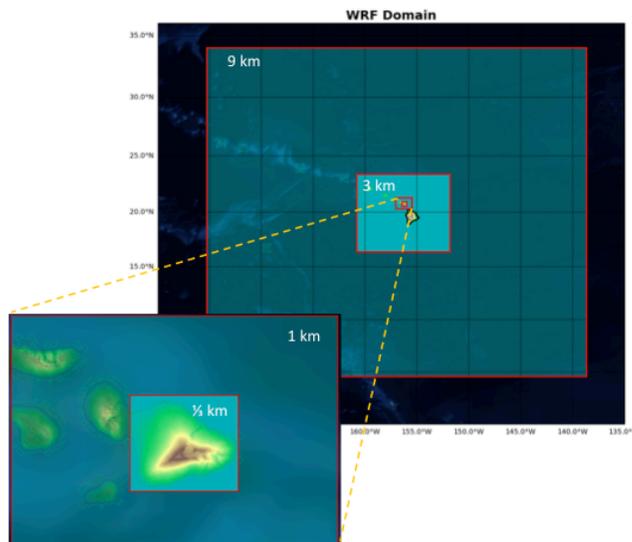


Fig. 2. WRF nested grid setup featuring an ultra-high resolution grid centered on the summit of Haleakala.

The WRF model is used with data assimilation in a constant configuration to produce daily 48-hour forecasts from 10/1/2016 to present. WRF is initialized with output from the Global Forecasting System (GFS) produced by the National Weather Service at 0.25° resolution. The WRF three-dimensional variational (3DVAR) data assimilation (DA) system is used to improve the quality and resolution of the initial conditions first prescribed by the global model. The DA system ingests local surface observations and aircraft reports of temperature, pressure, winds and moisture (also known as the Standard Meteorological Variables, SMV), UA soundings, and observed radiances from several satellites. WRF is initialized at 12 UTC with the output from a previous cold-start WRF simulation initialized with GFS data at 00 UTC and boundary conditions generated by WRFDA. WRF output is saved at 1-hour temporal resolution for domains 1 and 2, and 10-minute temporal resolution for domains 3 and 4. This process is illustrated in Fig. 3.

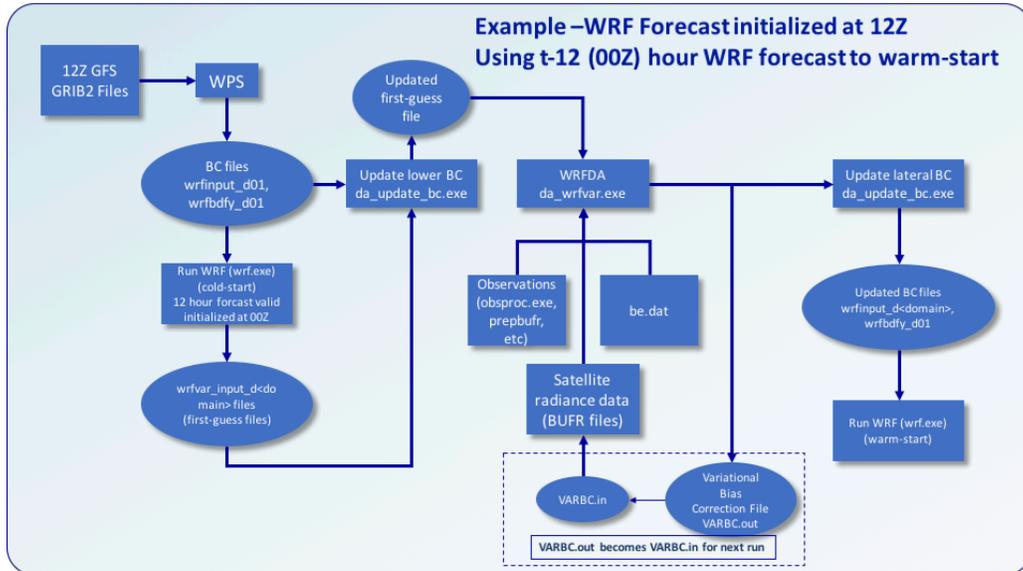


Fig. 3. WRFDA warm start configuration.

The WRF simulations use the physics options shown in Table 1 and were selected based on personal communication with a group at the University of Hawaii Department of Meteorology who have considerable experience performing WRF simulations over this same area (Steven Businger and Tiziana Cherubini, University of Hawaii at Manoa, ongoing communications). All parameterization options are the same for all domains, except for the cumulus parameterization which is only activated for the outermost 9km domain. The radiation scheme is called every 10 minutes. The performance of this WRF model configuration is discussed in [7].

Table 1. WRF physics parameterization schemes.

Parameterization Option	Scheme selection
Microphysics	WSM6
Longwave radiation	RRTMG
Shortwave radiation	RRTMG
Surface Layer	Monin-Obukhov
Land surface Model	Unified Noah
Planetary Boundary Layer	Mellor-Yamada-Janjic
Cumulus (d01 only)	Kain-Fritsch

b. *WRF-derived cloud layer classification algorithm*

In order to evaluate the ability of the WRF model to predict cloud cover over Haleakala, the standard WRF output variable CLDFRA is used. CLDFRA which is derived from [8] represents the cloud coverage/fraction of each grid cell at every vertical model level and has values between 0.0 and 1.0. A logical cloud presence/absence

variable is calculated based on whether the maximum CLDFRA value in the following atmospheric layers is greater than 0.05:

$$\begin{cases} \text{LowCloud: Surface} > \text{Pressure} \geq 690\text{mb} \\ \text{MidCloud: } 690\text{mb} > \text{Pressure} \geq 400\text{mb} \\ \text{HighCloud: } 400\text{mb} > \text{Pressure} \geq 175\text{mb} \end{cases}$$

In general, the configuration of WRF used here tends to exhibit a clear bias locally at the grid cell of Haleakala at all resolutions as compared to the validation data set described in the following subsection. This underestimation of cloud cover is an additional motivation for developing a hybrid RF-WRF cloud forecasting system.

c. Validation data

Whole Sky Imager (WSI) data taken from the summit of Haleakala is used to train and validate the WRF and Machine Learning model predictions over the timeframe October 2016 through July 2017. The WSI is a very useful tool because it takes high resolution digital images of the full sky dome down to the horizon during all times of the day – an example of a WSI image is displayed in Fig. 4. The WSI images are manually inspected for clouds at 10 minute intervals and assigned a cloud classification of clear, low clouds present, or mid/high clouds present so that they can be used for validation purposes in our Machine Learning approach. Due to the fact that this dataset is based on manual classifications of 2D images, there is no way to distinguish the presence of mid/high clouds on top of overcast low clouds. It is worth noting that although the visual inspection of images is subjective, the classification is performed with expert guidance and the resulting dataset exhibits a realistic diurnal cycle of clouds over Haleakala (Fig. 5). A maximum in cloudiness is observed in the early afternoon when trade wind inversion layer breaches the summit allowing for low clouds to creep over the summit. At night a minimum in clouds is observed consistent with the trade wind inversion lowering below the summit in the absence of heating from the sun.



Fig. 4. Visible WSI image of the entire sky dome on July 22nd, 2017.

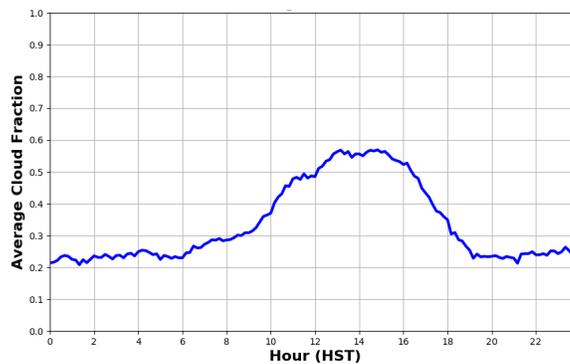


Fig. 5. Diurnal variation in WSI low cloud cover over Haleakala, HI over the time period 10/1/2017 – 7/31/2017.

3. METHODS

a) Random Forest theory and background

This study uses Machine Learning (ML) to improve WRF forecasts of cloud cover over Haleakala. ML algorithms are extremely powerful because they do not require a comprehensive understanding of physical processes. ML techniques instead use advanced computer science and statistical tools to train models that have a high predictive capacity. Nonetheless, it is still important to utilize ML tools in a scientifically meaningful way in order to reap their full benefits. This study takes advantage of the vast amounts of NWP data and corresponding validation data that have been cataloged over the last year for the purpose of forecasting cloud cover over Haleakala. For this study, ML is applied to WRF model output in order to determine important predictors and provide a benchmark for potential improvements over the native WRF output.

The ML technique demonstrated here is an ensemble learning method called Random Forest [9]. Random Forest (RF) is a powerful non-linear statistical analysis approach that has been shown to offer advantages in many different environmental realms. RF has proven applicable to a variety of problems from predicting atmospheric turbulence [10] to determining flood hazard risk [11]. RFs are ensembles of weakly-correlated and weak-learning decision trees that each vote on a single outcome. The overall result of a RF model can be either a binary classification of an outcome based on the majority vote from all the trees in the forest or a probability of an outcome based on the distribution of the votes within the forest. For the purposes of this project, the former binary classification approach is implemented.

For each tree in the random forest, a random subset (usually about 2/3) of the total number of samples is drawn with replacement from a “training” dataset that consists of predictor variables and associated “truth” values. The remaining portion (usually about 1/3) of the total number of samples is used to evaluate the unbiased mean square error (a.k.a. out-of-bag, OOB error) of the resulting decision tree model. At each node of the tree, a random *mtry* number of the input predictor variables are chosen which provides candidates for splitting the data. The data is split into two parts based on a condition/value of a single predictor variable (a.k.a. “feature”) such that similar samples end up in the same set. The measure based on which the locally optimal condition is chosen is called the Gini impurity index, which indicates how often a randomly chose sample from a set of data would be incorrectly labeled if it were labeled according to the distribution of labels within the dataset. The degree to which each feature decreases the weighted Gini impurity among the various trees in the forest (a.k.a. “feature importance”) is recorded through this process that is visualized in Fig. 6.

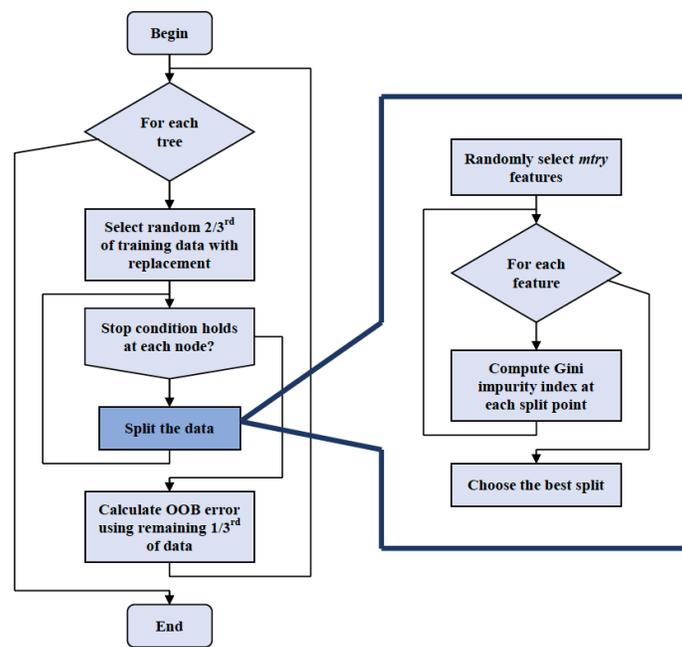


Fig. 6. Random Forest flow chart describing bagging and node split approach for each tree in the forest.

Due to the capability of RF to learn on the most important features, one might expect that adding more predictors to the RF model is always advisable in order to open up the possibilities for selecting features that result in low-impurity data splits. However, adding predictors indiscriminately can actually increase cross-correlations between variables, and potentially reduce the ability of the RF to learn on the most meaningful features. Therefore, it is important to choose predictor variables strategically and remove redundant and low-importance variables. Examination of the feature importances is iteratively performed here in order to identify the most useful predictors and avoid heavily correlated variables.

The RF approach is sensitive to the number of trees in the forest (*ntrees*) as well as the number of variables used to split the data at each tree node (*mtry*) [12,13]. For example, Fig. 7 shows the OOB error rate for various *ntrees* values given different *mtry* values for the local-surface experiment (described in Section 3e) over the time period 09/25/2016 – 05/31/2017. In general, the optimal *mtry* value is around the square root or one-third of the total number of predictors. Therefore, all RF experiments in this paper are performed with those 2 *mtry* values, unless the square root and one-third are approximately equal (which is often the case for experiments with very few predictor variables), in which case only one *mtry* value is evaluated.

Fig. 7 also demonstrates that increasing the number of trees generally results in more robust predictions, but only up to a point. While the spread between different *mtry* values is slightly smaller when using 1000 trees vs. 100 trees, the overall OOB Error rate is about the same. Due to the fact that training more trees requires significantly more computational time, the optimal *ntrees* value is here considered to be around 100. These results are consistent across the various experiments discussed in this paper. Hence, only the results for RFs with *ntrees*=100 are reported in this paper.

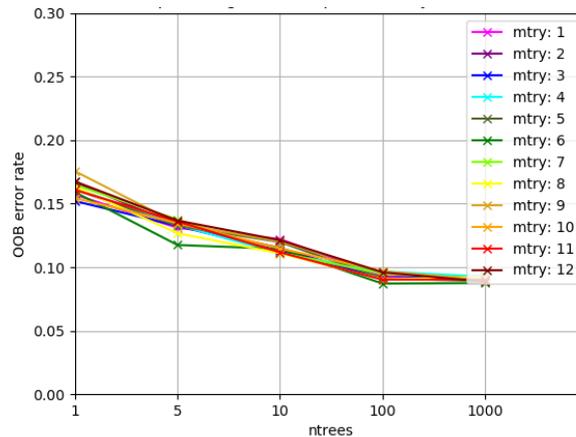


Fig. 7. Out-of-bag (OOB) error rate for several RF experiments which vary the number of trees (*ntrees*) and the number of randomly selected predictor variables which are used for determining data splits at each tree node (*mtry*) for RFs that use WRF local surface predictors over the time period 9/25/2016 – 5/31/2017.

b) Low and Mid/High Cloud separation

The overall goal of this project is to produce a column-integrated cloud decision for the atmosphere directly overlying the Haleakala summit in order to determine potential interference along the line of sight to a satellite. In order to achieve this goal, RF predictions based on WRF model output variables are generated for low cloud cover and for mid/high cloud cover separately and then combined into a single column cloudiness decision. The reason for the separation of low and mid/high cloud predictions is twofold. First, the nature of the validation dataset is somewhat exclusive to the type of clouds present (i.e. high clouds are only identifiable when there are no low clouds present), and therefore it is difficult to create a meaningful column-integrated cloudiness decision validation product. Second, the meteorological dynamics responsible for low cloud and mid/high cloud formation are often different and quasi-independent. As such, the meteorological variables that are most useful in training a RF model are likely to be different for the different cloud layers. Therefore, one might expect that the accuracy of the overall prediction will be improved with this two-pronged approach to predicting overall cloudiness. This paper describes only the first half of this approach – the prediction of low clouds. That is, the WRF-based RFs produce predictions of the presence (1) or absence (0) of low clouds directly overlying the Haleakala summit.

c) Experiment protocol

The WRF predictor and WSI validation datasets span the time period 10/1/2016 – 07/31/2017. RF models are trained on odd numbered days (training dataset) and evaluated on even numbered days (testing dataset) during this time period. This data sampling procedure was chosen after trial and error with other approaches and was ultimately selected because it ensures adequate sampling of the event of interest (i.e. cloudy conditions). The approach also allows for sampling of the broad range of possible meteorological conditions across the interannual spectrum. Although the WRF model output is saved at 10-minute resolution, only model data at the top of each hour is used to make a prediction of low cloud presence. Each WRF-based RF prediction is validated against the manual low cloud classifications from WSI for each corresponding hour. In order for a sample (i.e. data at a particular hour) to be included in the training or testing datasets, all of the input variables must contain real values; that is, samples containing any missing data are not included.

d. Predictors

Presently, only WRF variables from the inner most domain at 1/3 km resolution are used, but the predictor set may eventually be expanded to include variables from the coarser domains as deemed useful. In addition, only WRF variables from the forecast matching the target time are utilized here in the training/evaluation the RF algorithm (e.g. WRF forecast output for 0400 is used to predict low cloud cover at 0400). A preliminary examination of the sensitivity of the WRF-based RF approach to offset forecast inputs at various temporal lags (e.g. using the 0300 WRF forecast output to predict low cloud cover at 0400), demonstrated that there was no improvement over using the temporally corresponding WRF forecast output.

Since WRF is initiated daily at 12Z and run for 48-hours, there are two values for each variable available each day – one from the current day’s forecast and one from the prior day’s forecast. The RF algorithm ingests both the current and the prior days’ forecast variables in order to maximize the likelihood of using the most accurate forecast data. It is generally the case that both forecasts are useful to the RF model.

Several key WRF variables – both native and derived – were selected as input variables to the RF algorithm based on their importance in the formation and/or advection of clouds as well as their high feature importance ranks determined through various RF sensitivity tests. The selected WRF variables fall into the three categories described below, which represent differing natures and spatial extents. It is worth noting that a fourth category of predictors – 3D grid-cell predictors where variables of interest were ingested by the RF algorithm at every grid cell within a certain area around the Haleakala summit – was evaluated for a shorter time period. However, this type of information did not produce accurate results and was prohibitively computationally intensive. This is perhaps to be expected because using grid-cell level predictors produces high cross-correlation between predictors and also results in a vast set of predictors, both of which can degrade the overall performance of ML algorithms [2].

1) Local surface predictors

The variables 2-meter temperature, 2-meter relative humidity, 10-meter u-wind, 10-meter v-wind, surface pressure, 10-meter wind direction, and 10-meter wind speed locally at the WRF grid cell which encompasses the Haleakala peak were selected to evaluate how well the RF performs when provided local WRF forecast data near the surface. These 7 variables all have a role in cloud formation locally, with 2-meter relative humidity exhibiting the highest relative feature importance among them. When incorporating both the current day’s forecast output and the prior day’s forecast output, this experiment family consists of 14 total predictors.

2) Regional 3D moisture predictors

Three-dimensional WRF model output is utilized in order to inform the RF algorithm about the larger-scale dynamics of the island of Maui, which can often be influential in local cloud presence. Through a series of tests, it was determined that regional moisture variables in and around the level of the Haleakala summit have strong predictive capacity for local low cloud cover. Hence, regional statistics of three moisture variables – relative humidity, dew point temperature and dew point depression – on key pressure levels were selected as input variables to the RF. In order to capture the larger-scale atmospheric dynamics, the inner WRF domain was divided into 4 quadrants with the Haleakala grid cell at the center such that Q1 is the NE quadrant, Q2 is the SE quadrant, Q3 is the SW quadrant and Q4 is the NW quadrant. The average relative humidity, maximum relative humidity, average dew point temperature, maximum dew point temperature, average dew point depression and minimum dew point depression were calculated for all 4 quadrants on the following pressure levels: 775, 750, 730, 720, 710, and 700

mb. This regional 3D moisture information is provided to the RF as a set of 72 predictor variables. Note that RF experiments using regional 3D moisture variables on an expanded set of pressure levels from 1000mb to 500mb did not demonstrate a significant difference in the overall results. When incorporating the current day's forecast output, the prior day's forecast output, the 6 pressure levels, and the 4 quadrants, this experiment family consists of 288 total predictors.

3) Regional inversion predictors

As previously discussed, the height of the trade-wind inversion is important in determining how high convective clouds can grow. To test the predictive potential of the large-scale inversion within the framework of the RF-WRF model, an inversion is identified at each grid cell in the WRF inner domain based on where the vertical potential temperature change exceeds 0.05K as in [14]. Both the bottom and top of the first identifiable inversion are recorded for each vertical column. A series of sensitivity experiments reveal that several regional WRF inversion variables exhibit high predictive capacity in the RF model. Specifically, the average of the bottom height of the inversion taken over each of the four quadrants, as described in the previous subsection, is chosen because it exhibits strong predictive capacity. In addition, the average of the top height of the inversion computed only for the WRF grid cells for which the terrain height is less than 1500m is chosen as a potentially powerful predictor. This derived parameter – hereafter referred to as the regional terrain-masked inversion top height – describes the height of the inversion only in the areas not directly influenced by local terrain. When incorporating both the current day's forecast output and the prior day's forecast output, this experiment family consists of 10 total predictors.

4. RESULTS

In order to measure the quality of the WRF-based RF predictions and compare them to the performance of WRF's native output, several skill scores for binary classification problems (e.g. no-yes, 0-1 or in our case: clear-cloudy) are reported. These skill scores are based on Table 2 and Equations 1-4 as in [15].

Table 2. Contingency table representing the performance of a forecast compared to observations for binary-type events (e.g. yes/no, present/absent, 0/1 etc.).

2x2 Contingency Table		Event Observation	
		Yes	No
Event Forecast	Yes	A	B
	No	C	D

$$\text{Skill score equations: } \left\{ \begin{array}{l}
 \text{POD} = \frac{A}{A + C} \quad (1) \\
 \text{FAR} = \frac{B}{A + B} \quad (2) \\
 \text{bias} = \frac{A + B}{A + C} \quad (3) \\
 \text{CSI} = \frac{A}{A + B + C} \quad (4)
 \end{array} \right.$$

The Probability of Detection (POD) represents how many times the forecast correctly predicts the occurrence of the event (i.e. forecasted to be cloudy when it is indeed cloudy) divided by the total number of times the event actually occurs. The False Alarm Ratio (FAR) represents how many times the forecast incorrectly predicts the occurrence of the event (i.e. forecasted to be cloudy but it is actually clear) divided by the total number of times that the forecast predicts the occurrence of the event. The bias indicates the number of times the event is forecasted to occur divided by the number of times the event actually occurs. The Critical Score Index (CSI) – also known as the threat score – represents the fraction of observed and/or forecasted events that are correctly predicted. These skill scores are designed to evaluate the ability of a model to predict rare events – in this case low clouds.

The previously described skill scores can be visualized on a performance diagram as demonstrated by [15] where POD is on the y-axis and 1-FAR (called the Success Ratio [SR]) is on the x-axis. Dashed lines on the Roebber performance diagram represent the bias and hyperbolic lines represent the CSI. Forecasts are deemed better the closer they are to the top right corner of the plot (higher POD, CSI and SR, and bias approaching 1).

Fig. 8 displays the results of several WRF-based RF experiments that use the three different types of WRF input variables (local surface predictors, regional 3D moisture predictors and regional inversion predictors). These are shown as individual experiments as well as an experiment with all three types of predictors in combination for the time period October 2016 through July 2017. The WRF CLDFRA logical low cloud variable described in Section 2 is also displayed on Fig. 8. As stated earlier, each RF experiment is reported for $n_{trees}=100$ and four different m_{try} values. Note that for experiments with very few predictors, there may only be three m_{try} values evaluated because the rounded values of the square root and one-third are equal. Table 3 displays the mean skill scores across the three or four m_{try} values for each RF experiment as well as the skill scores for the WRF CLDFRA variable. As can be seen in Fig. 8, there is not a large variation between different m_{try} values within a given $n_{trees}=100$ experiment family.

It is clear from Fig. 8 and Table 3 that RFs trained on high-resolution WRF model output significantly outperform the native WRF CLDFRA output, with all RFs having higher POD and CSI scores and biases closer to 1.0. Although WRF CLDFRA exhibits a very low FAR (high SR), it has a low POD and a bias score that is much less than one. In essence, although the WRF model rarely predicts low clouds when there are not actually any low clouds present, it tends to under-predict the occurrence of low clouds overall. For the purposes of anticipating atmospheric interference in an environment where low cloud cover is the minority condition, having a low FAR score alone is insufficient, and needs to be coupled with high POD, high CSI and a near-1 bias, which is what the WRF-based RF models achieve.

Performance Diagram Summarizing the SR, POD, bias, and CSI

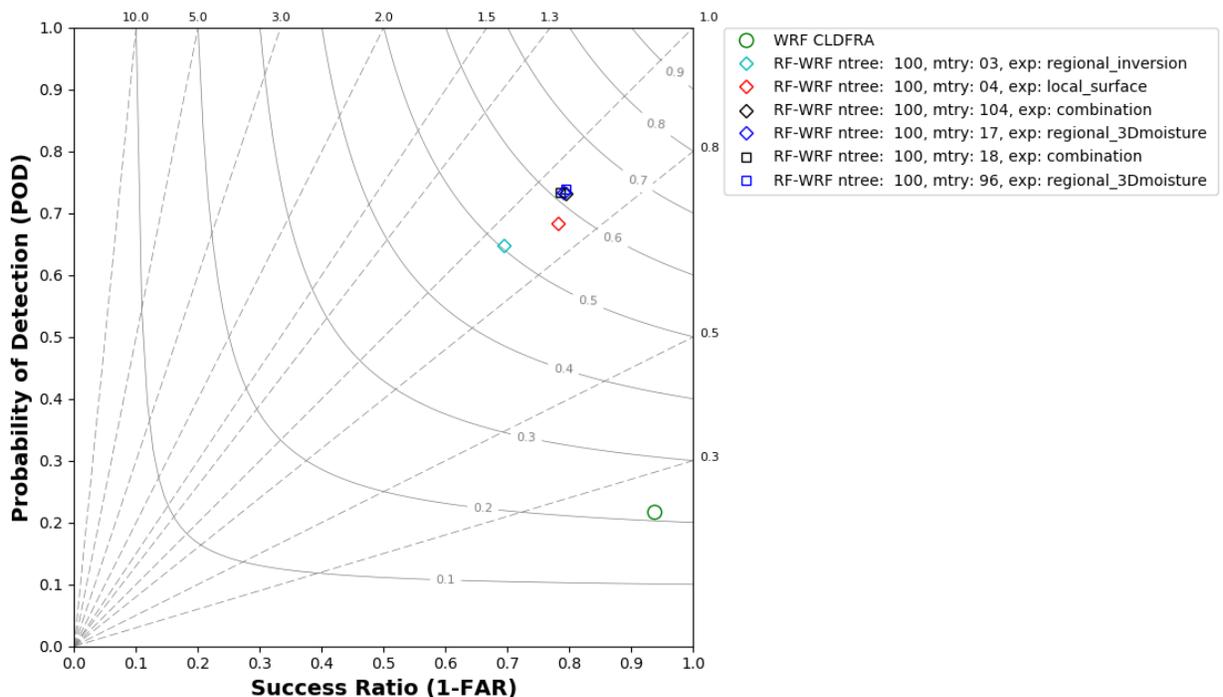


Fig. 8. Performance Diagram of WRF-based RF experiments performed over the time period 10/1/2016 – 7/31/2017. RF-WRF experiments are trained on approximately 2690 samples (odd days) and evaluated on approximately 2640 samples (even days) depending on the availability of the various types of WRF predictors and WSI validation data.

Table 3. Average skill scores across the one or two *mtry* values for each unique WRF-based RF experiment family, and skill scores for the WRF CLDFRA native output over the time period 10/1/2016 – 7/31/2017.

Product	Average POD	Average FAR	Average Bias	Average CSI
RF-WRF using local surface predictors	0.683	0.218	0.874	0.574
RF-WRF using regional 3D moisture predictors	0.736	0.208	0.929	0.617
RF-WRF using regional inversion predictors	0.648	0.307	0.935	0.503
RF-WRF combination of all predictors	0.733	0.211	0.929	0.612
WRF CLDFRA native output	0.218	0.063	0.233	0.215

An inter-comparison of the different WRF-based RF models reveals that RFs that are trained on regional 3D moisture predictors perform best in terms of both the POD and SR. The fact that the local surface predictors do not result in the best predictions is perhaps to be expected because it is often the case that it is much more difficult for NWP models to predict the presence of clouds locally than to predict the qualitative spatial pattern of cloud cover overall. In addition, RFs which use a combination of local surface, regional inversion and regional 3D moisture predictors do not perform notably better than those which use only regional 3D moisture predictors. This is reflected in the relative feature importances, an example of which is given in Fig. 9. From Fig. 9 one can see that the quadrant minimums of dew point depression and the quadrant maximums of relative humidity are the most influential in determining low-impurity splits at the tree nodes. Interestingly, the most predictive variables are those that represent moisture at pressure levels slightly below the summit. The summit of Haleakala is at about 710mb, yet the relative humidity at 750mb is the strongest predictor when it comes to forecasting low cloud cover over the summit.

Although the quadrant averages of the bottom height of the inversion are relatively important, their addition to the combination experiment does not appear to improve the overall performance of the RF-WRF model overall. This could indicate that regional inversion characteristics are poorly simulated by WRF or that the inversion identification algorithm employed here is suboptimal. This could also be due to fact that the inversion predictors are highly correlated with regional 3D moisture predictors because the location of the large-scale inversion is a first order control on the amount of moisture in the upper levels of the atmosphere. It is notable as well that the terrain-masked regional inversion top height variable does not show up in the top 40 predictors, implying that the quadrant averages of the bottom height of the inversion have more predictive capacity. The local surface predictors do have some impact on tree split nodes in the combination experiments, but they are about an order of magnitude less important than the top predictors. Hence, WRF-based RFs that are built upon highly descriptive regional predictors can perform just as well as RFs that are based on many different types of predictors in combination.

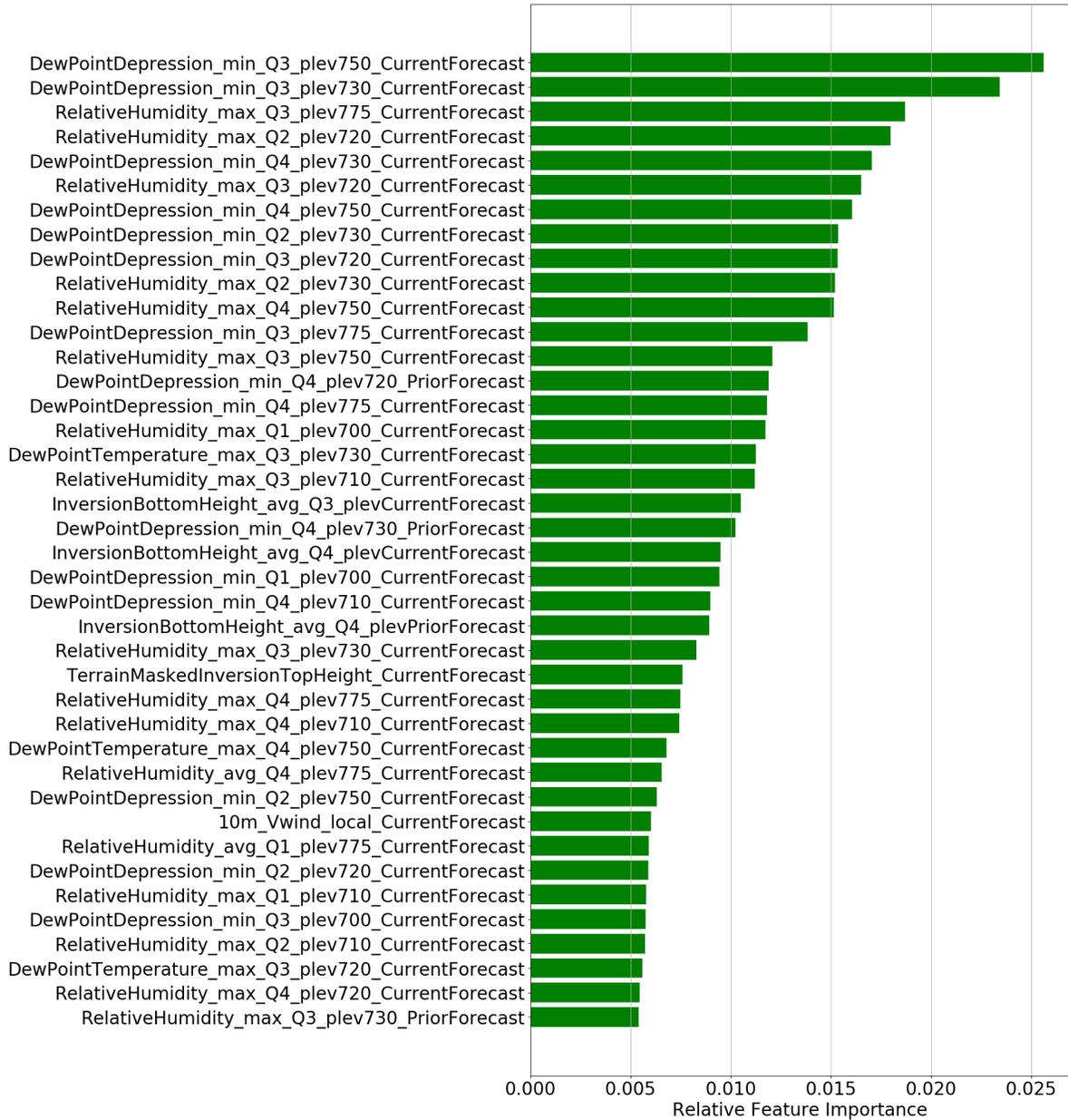


Fig. 9. Relative feature importances of the top 40 predictors (out of 312 total predictors) for the best-performing WRF-based RF experiment using a combination of local surface, regional 3D moisture and regional inversion predictors with $n_{trees} = 100$ and $m_{try} = 104$. The number following “_Q” indicates the quadrant number. And the numbers following “_plev” indicate the pressure level in mb.

The performance of the WRF-based RF algorithm is heavily dependent on the time periods over which the algorithm is trained and evaluated. In fact, when the 10-month time period is broken down into two 5-month periods Oct 2016 – Feb 2017 and Mar 2017 – Jul 2017, the WRF-based RF model that is trained and evaluated on the former time period (Table 4) produces PODs upwards of 0.78 and significantly outperforms the WRF-based RF model that is trained and evaluated on the latter time period (Table 5). This may be because the dynamics responsible for local low cloud formation are more accurately simulated by WRF during some seasons. In fact, the WRF CLDFRA forecasts have lower POD during the Mar 2017 – Jul 2017 time period. While the number of samples utilized was similar between these two time periods, it is notable that there were many more cloudy events during the latter time period. It is important to be aware of these biases when producing cloud cover forecasts during different times of the year.

Table 4. Same as Table 3 except for the time period 10/1/2016 – 2/28/2017.

Product	Average POD	Average FAR	Average Bias	Average CSI
RF-WRF using local surface predictors	0.710	0.197	0.885	0.605
RF-WRF using regional 3D moisture predictors	0.772	0.212	0.979	0.639
RF-WRF using regional inversion predictors	0.710	0.292	1.003	0.550
RF-WRF combination of all predictors	0.778	0.214	0.990	0.642
WRF CLDFRA native output	0.244	0.050	0.256	0.241

Table 5. Same as Table 3 except for the time period 3/1/2017 – 7/31/2017.

Product	Average POD	Average FAR	Average Bias	Average CSI
RF-WRF using local surface predictors	0.638	0.246	0.846	0.528
RF-WRF using regional 3D moisture predictors	0.691	0.221	0.888	0.577
RF-WRF using regional inversion predictors	0.616	0.362	0.966	0.457
RF-WRF combination of all predictors	0.714	0.200	0.892	0.606
WRF CLDFRA native output	0.169	0.092	0.186	0.166

5. CONCLUSION AND DISCUSSION

This study demonstrates a significant improvement upon the local WRF low cloud cover predictions at the Haleakala summit using a Random Forest approach trained on and validated with in-situ data. The results of the WRF-based RF experiments reported here show substantial gains in the POD and CSI while reducing the overall bias compared to the native WRF CLDFRA output. This work exhibits the potential that Machine Learning techniques have to improve NWP forecasts. Such tools can be used to mitigate the risk of atmospheric fading on optical communications systems, thereby improving the accuracy of link handover decisions.

In particular, this study reveals that WRF-based RF models perform best when provided with regional 3D information based on fields that are relevant to the atmospheric phenomena of interest. In this case 3D moisture variables such as relative humidity on select pressure levels are key to predicting low cloud cover locally over Haleakala. This study also reveals that adding more variables and/or different types of variables (e.g. local surface predictors) does not necessarily improve the overall prediction of low cloud cover locally. The fact that adding more pressure levels to the regional 3D moisture predictor set did not significantly change the overall results, as mentioned in Section 3e, is also a testament to this conclusion.

It is important to understand that the RF results presented here are based on the assumption that the validation product – in this case WSI manual cloud classifications – is perfect. Although the WSI dataset exhibits a reasonable diurnal cycle, it is by no means perfect. There are, however, plans to develop a more accurate validation dataset in order to improve the RF predictions for imaging and FSOC applications. Nonetheless, the results shown here are highly promising and the fact that the RF produces such high skill scores indicates the model is learning based on a physically realistic validation dataset.

This work has thus far exclusively focused on predicting cloud cover locally – that is at a single point (e.g. the Haleakala summit). However, future work will expand the RF approach to other locations in order to develop a 2D map of cloud predictions. Given that WSI data is only available locally, the expanded version of the WRF-based RF algorithm will be validated against cloud classifications derived from satellite-determined cloud top heights.

Future efforts will attempt to improve the predictive accuracy of the current low cloud model by experimenting with seasonal models, the addition of other high-resolution WRF variables and derived quantities, and the addition of lower-resolution WRF output from the coarser domains. The next phase of this work is to develop a WRF-based RF model for high cloud prediction, which will be validated using a combination of the WSI manual cloud classification dataset and satellite-derived cloud classifications based on cloud top heights. Although the most influential predictors may end up being different from those used in the low cloud model, the process for identifying useful WRF variables will be similar. This process involves a great deal of tuning with regards to the data sampling protocol and the selection of input variables – factors which can be heavily influential in the resulting accuracy of the RF predictions. Although these choices must be made somewhat subjectively (it is impossible to examine every possible combination of input variables), expert knowledge is leveraged at each stage of the model tuning to determine the most highly influential variables. Ultimately, the best low cloud RF model and the best high cloud RF model will be used in combination to make a final column-integrated decision that will represent the likelihood of atmospheric interference due to cloud cover over Haleakala.

A significant challenge to using NWP model output as input to a Machine Learning algorithm such as Random Forest is how to take advantage of the spatially and temporally varying data in a scientifically meaningful way. While regional averages, maximums, and minimums have been shown to capture some of the large-scale dynamics within the innermost WRF domain, the spatial evolution of clouds over time is very dynamic. For this reason, it is extremely difficult to quantify the large-scale atmospheric evolution with a single or even multiple variables, no matter how intelligently or empirically designed. Hence, there are efforts underway to develop methods of including spatially and temporally varying information into a RF-type approach.

Acknowledgements

The Python scikit-learn [16] sklearn module implementation of Random Forest was used to perform the suite of WRF-based RF experiments reported in this paper. The authors would like to thank Michael Mason for writing the foundational RF code, Mary Ellen Craddock for testing the code with a sample dataset, and Heather Kiley for maintaining the WRF data and data servers, all of Northrop Grumman Corporation. In addition, the authors would like to thank Oliver Alliss of Virginia Tech for providing the manual classifications of the whole sky imager truth dataset.

References

- [1] Kwong, K.M., M. Wong and J. Liu, An Artificial Neural Network with Chaotic Oscillator for Wind Shear Alerting, *Journal of Atmospheric and Oceanic Technology*, Vol. 29, 1518–1531, 2012.
- [2] Aler, R., R. Marin, J.M. Valls and I.M. Galvan, A Study of Machine Learning Techniques for Daily Solar Energy Forecasting Using Numerical Weather Models, In: Camacho D., L. Braubach, S. Venticinque, C. Badica (eds) *Intelligent Distributed Computing VIII. Studies in Computational Intelligence*, Vol. 570, 269–278, 2015.
- [3] Collins, W. and P. Tisot, An artificial neural network model to predict thunderstorms within 400km² South Texas domains, *Meteorological Applications*, Vol. 22, 650–665, 2015.
- [4] Mecikalaski, J.R., J.K. Williams, C.P. Jewett, D. Ahijevych, A. LeRoy and J.R. Walker, Probabilistic 0–1-h Convective Initiation Nowcasts that Combine Geostationary Satellite Observations and Numerical Weather Prediction Model Data, *Journal of Applied Meteorology and Climatology*, Vol. 54, 139–159, 2015.
- [5] Gala, Y., A. Fernandez, J. Diaz and J.R. Dorransoro, Hybrid machine learning forecasting of solar radiation values, *Neurocomputing*, Vol 176, 48–59, 2016.

- [6] Skamarock, W.C., J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, M.G. Duda, X.-Y. Huang, W. Wang and J.G. Powers, A Description of the Advanced Research WRF Version 3. *NCAR Tech. Note NCAR/TN-475+STR*, 113 pp., 2008.
- [7] Alliss, R.J., B.D. Felton, M.E. Craddock, H. Kiley and M. Mason, 21st Century Atmospheric Forecasting for Space Based Applications, *Advanced Maui Optical and Space Surveillance Technologies Conference*, 2016.
- [8] Xu, K.-M. and D.A. Randall, A semiempirical cloudiness parameterization for use in climate models, *Journal of Atmospheric Science*, Vol. 53, 3084–3102, 1996.
- [9] Breiman, L., Random forests. *Machine Learning*, Vol. 45, 5–32, 2001.
- [10] Williams, J.K., J. Craig, A. Cotter and J.K. Wolff, A hybrid machine learning and fuzzy logic approach to CIT diagnostic development. *AMS Fifth Conference on Artificial Intelligence Applications to Environmental Science*, 2007.
- [11] Wang, Z., C. Lai, Z. Chen, B. Yang S. Zhao and X. Bai, Flood Hazard risk assessment model based on random forest, *Journal of Hydrology*, Vol. 527, 2015.
- [12] Cutler, D.R., T.C. Edwards Jr., K.H. Beard, A. Cutler, K.T. Hess, J. Gibson and J.J. Lawler, Random Forests for Classification in Ecology, *Ecology*, Vol. 88: 2783–2792, 2007.
- [13] Strobl, S., A.-L. Boulesteix, T. Kneib, T. Augustin and A. Zeileis, Conditional Variable Importance for Random Forests, *BMC Bioinformatics*, Vol. 9, 11 pp., 2008.
- [14] Heffter JL., Transport Layer Depth Calculations, *Second Joint Conference on Applications of Air Pollution Meteorology*, New Orleans, Louisiana, 1980.
- [15] Roebber, P.J., Visualizing Multiple Measure of Forecast Quality, *Weather and Forecasting*, Vol. 24, 601–608, 2009.
- [16] Pedregosa, F., et. al, Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, Vol. 12, 2825-2830, 2011.