Accelerated AI Powered Atmospheric Predictions for Space Domain Awareness Applications

Danny Felton
Northrop Grumman Corporation
Mary Ellen Craddock, Heather Kiley, Randall J. Alliss, Eric Page, Duane Apling
Northrop Grumman Corporation

ABSTRACT

Space based laser and surveillance applications are often impacted by atmospheric effects. Atmospheric attenuation and distortion caused by aerosols, clouds and optical turbulence can produce harmful effects thereby negatively impacting mission outcomes. A paper briefed at the 2019 AMOS conference described ground-based instrumentation installed at the Haleakala summit in 2017. Still actively collecting data, these instruments are providing unprecedented real-time characterization of the space environment, including minute to minute atmospheric transmission losses. Although real-time measurements are a first step for understanding and characterizing the space environment, they are not alone sufficient. Accurate, short-term atmospheric predictions of the space environment are necessary for many applications in order to optimize mission planning. Although atmospheric predictions are nothing new, their skill has recently been greatly improved with the use of 21st century Artificial Intelligence (AI) technologies. These technologies are a union between high performance computing (HPC) and Deep Learning (DL). The ability to train prediction models with terabytes of data from ground-based atmospheric collection systems and accelerate both their training and inferencing with the use of Graphical Processing Units (GPUs) is the subject of this presentation.

This study focuses on three time scales of prediction. These timescales include short-range (0 to 60 minutes), mid-range (1 hour to 3 hours), and long-range (3 to 48 hours). These time scales represent various decision points for laser and/or surveillance applications and missions. In the short-term prediction case, several DL techniques are applied to local data collected from an optical ground station (OGS). These DL techniques include the use of a U-Net convolutional neural network and an ensemble of multi-layer perceptron (MLP) and Random Forest (RF) models. The MLP is used for point data collected from instruments like a laser ceilometer and Infrared Cloud Imager (ICI). For the intermediate time scale, both a convolutional Long Short Term Memory (LSTM) network and a U-Net are trained with imagery from a collection of NOAA geostationary satellite images of clouds. Finally, a combined U-Net and an autoencoder neural network are used to train atmospheric predictors simulated from an HPC Numerical Weather Prediction (NWP) model to make long-range predictions. The NWP produces many terabytes of data and, therefore, the use of these neural networks is ideal to optimize its predictive ability. Several HPC resources are utilized for this study. These include an in-house GPU node consisting of four NVIDIA Tesla V100 GPUs as well as resources at the Maui High Performance Computing Center (MHPCC). Results indicate that in nearly all cases these prediction technologies are outperforming persistence with very little bias. The ability to make predictions in real-time using HPC and DL inferencing is now the focus moving forward and will be reported on at the conference.

1. INTRODUCTION

Atmospheric attenuation and distortion reduce the efficacy of space-based laser and surveillance applications. In particular, clouds can partially or fully obscure targets, and block or require a reduction of the data rate of optical communications systems. However, with accurate characterization and prediction of atmospheric effects, many of the negative impacts can be mitigated. The objective of this research is to develop and mature a state-of-the-art atmospheric prediction system that generates high-resolution predictions of atmospheric attenuation to support decision aids for space-based laser and surveillance applications. To accomplish this, advances in HPC and AI are combined with terabytes of high-resolution ground- and space-based atmospheric data collections. Several HPC resources are used to process the ground and satellite-based data required for this study, and the training and inferencing of the AI prediction technologies is accelerated using four NVIDIA Tesla V100 GPUs. This technology is used to make multi-time scale atmospheric predictions: out to 1 hour; out to 2+ hours; and out 48 hours.

The requirements for atmospheric predictions at these three time scales are quite different, and, therefore, necessitate different sources of data, computing resources, and prediction techniques. Short-range predictions of a few minutes to an hour are most accurate when very high-resolution, local data can be used. Therefore, for predictions out to 1 hour, a ground-based infrared cloud imager (ICI) and co-located laser ceilometer provide the data, and an ensemble of a Random Forest (RF) and multi-layer perceptron (MLP) model is trained to make these predictions. For mid-range predictions out to a few hours, the ground-based imager is insufficient since clouds that may impact the system in 1-3 hours are likely not yet in the field of view of the local ICI. Therefore, cloud products from NOAA geostationary satellites is used for these mid-range predictions. For this effort, we developed and tested both a Long Short Term Memory (LSTM) network and a U-Net convolutional neural network trained and validated on satellite-derived cloud masks. The U-Net proved to substantially outperform the LSTM, and only the results of the U-Net will be presented. Accurate long-range cloud predictions cannot be derived from ground or satellite based cloud evolution. Therefore, Numerical Weather Prediction (NWP) is used as the basis for predictions out to 48 hours or more. The usefulness and limitations of NWP predictions are well-documented, and in particular, the clear bias of the high-resolution Weather Research and Forecasting (WRF) NWP model at Haleakala has been quantified [1, 2]. In this study, two DL techniques, a U-Net and an MLP, are investigated in order to improve the 48 hour cloud predictions of the WRF model. The results of these modeling efforts show that for all three time scales, the AI prediction technologies substantially outperform the baseline predictions.

The data and techniques used to create the components of a multi-time scale atmospheric prediction system are described in this paper. Section 2 describes the development of an MLP model to make short-range predictions of clouds based on cloud classifications from a ground-based ICI and laser ceilometer. The development of an improved satellite-derived cloud analysis using a U-Net convolution neural network, the U-Net Cloud Analysis (UCA) is presented in Section 3. The use of a U-Net to make mid-range cloud predictions based upon the UCA is discussed in Section 4. Finally, improvements in the long-range cloud prediction are described in Section 5.

1.1 INSTRUMENTATION AND DATA

A high-quality labeled dataset is required to train and validate any AI model. Several sources of cloud data are used in the training and validation of the AI models described in this work, the largest of which is a satellite-derived cloud classification product called the Cloud Mask Generator (CMG). Spanning more than 24 years, Northrop Grumman's CMG ingests Geostationary Operational Environmental Satellite (GOES 8-15) multispectral imagery (at 4 km, 15 minute resolution) and applies a series of single- and multi-spectral tests to detect clouds [3,4]. The GOES imager has 5 bands: visible (0.6 μm), shortwave infrared (3.9 μm) (SWIR), water vapor (6.7 μm), longwave infrared (10.7 μm) (LWIR), and split window (11.2 μm). The water vapor channel is not used for cloud detection and is replaced by a fog product at night and a shortwave reflectivity product during the day. The long wave infrared and split window channels are differenced to facilitate cirrus cloud detection. The CMG generates a cloud/no cloud decision for every pixel by computing the difference between the LWIR temperature, visible albedo, derived products, and the dynamically computed clear sky background (CSB) each time a new GOES image is available. The classification of a pixel as clear or cloudy is based on where the calculated difference falls with respect to the threshold confidence range. Threshold confidence ranges for each test are spatially and temporally defined. The limit of the CMG cloud detection is estimated to be between 1.0 and 1.5 dB [4].

A second source of high quality cloud data is a ground-based, fully calibrated, passive infrared instrument that provides nearly horizon to horizon coverage of the sky [5]. Deployed on the summit of Haleakala in the summer of 2017, the Infrared Cloud Imager (ICI) consists of a FLIR Photon 640 camera and electronics enclosure. The photon 640 comes mounted underneath a Stingray full sky lens. A hatch with a rain sensor protects the lens during inclement weather. The ICI produces calibrated sky radiances at each pixel within the skydome and has been collecting images at one minute resolution for more than four years. A cloud retrieval algorithm is used to interpret each image at the pixel level as cloud or no cloud. This algorithm uses the clear sky background (CSB) technique [4] which evaluates many sky radiance images as a function of time of day and identifies the 10^{th} percentile lowest values. Any deviation in sky radiance from the CSB is interpreted as a high confidence cloudy pixel.

A final source of labeled cloud data is the Vaisala Ceilometer CL51. The CL51 employs pulsed diode laser LIDAR technology, where short, powerful laser pulses are sent out in a vertical direction. The reflection of light, backscatter

– caused by haze, fog, mist, precipitation, and clouds – is measured as the laser pulses traverse the vertical column above the site. The resulting backscatter profile is processed and stored at six second intervals to compute cloud base heights and transmission loss. Backscatter profiles are available up to approximately 16 kilometers above ground level. This provides a characterization of cloud base height and thickness of up to three cloud layers, as well as an estimate of the transmission loss derived from the backscatter profile.

1.2 PERFORMANCE CHARACTERIZATION

Several skill scores for binary classification problems (e.g. no-yes, 0-1 or in this study: clear-cloudy) are computed to measure the quality of the cloud forecasts and compare them to the independent labeled cloud datasets. These skill scores are based on the contingency table and equations shown in Table 1 and Equations 1-4 [6], and will be used to assess the performance of each of the techniques described in this paper.

Table 1. 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 Prediction	Yes	A	B
	No	C	D

$$Skill\ Score\ Equations: \begin{cases} POD = \dfrac{A}{A+C} & (1) \\ FAR = \dfrac{B}{A+B} & (2) \\ bias = \dfrac{A+B}{A+C} & (3) \\ CSI = \dfrac{A}{A+B+C} & (4) \end{cases}$$

The Probability of Detection (POD) represents how many times the forecast correctly predicts the occurrence of the event (i.e. forecast 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. forecast to be cloudy but it is actually clear) divided by the total number of times the forecast predicts the occurrence of the event. The bias indicates the number of times the event is forecast to occur divided by the number of times the event actually occurs. The Critical Success 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 clouds. These skill scores can be visualized on a performance diagram 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, see Fig. 1). Scores to the lower right of the diagonal reflect a clear bias, while scores to the upper left of the diagonal reflect a cloudy bias.

2. ENSEMBLE MLP SHORT-RANGE CLOUD PREDICTION

The atmospherics above Haleakala have been well-characterized over the past several years [5]. As measurements from instrumentation on the Haleakala summit have shown, the Haleakala summit is cloud-free over 70% of the time. Therefore, a simple persistence forecast would work very well in most cases. By definition, however, persistence does not predict change and therefore, during times when clouds are moving *in and out* of a line-of-sight (LOS), this technique would fail. Fig. 1 (top left) shows a Roebber plot indicating the performance of a persistence forecast using clouds retrieved from the ICI as a function of forecast length. The data show that for longer forecasts, e.g. 15, 30, 60 minutes, a simple persistence forecast has a significant clear bias and the POD decreases as forecast length increases. Forecasts lengths of 1, 2, 3, 4, 5, 10, 15, 30 and 60 minutes are shown in Fig. *1*.

To improve upon a simple persistence cloud forecast, two separate ML models were developed to predict the probability of clouds out to 60 minutes from the current time. The first technique is an ensemble learning method called Random Forest (RF). RFs are ensembles of weakly-correlated and weak-learning decision trees that each vote on a single outcome. For each of the trees in the RF, a random subset of the total number of samples is drawn from a training dataset. At each node of the decision trees, the dataset is split into two parts based on a condition/value of a single predictor variable. 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. The second ML model is a multi-layer perceptron model (MLP). A perceptron is an algorithm used to perform binary classifications of cloudy or clear. It produces a single output based on inputs by forming a linear combination using input weights. An MLP is a deep artificial neural network composed of more than one perceptron. It contains an input layer that receives the signal, an output layer that makes the prediction, and hidden layers in between that do the computational work. The MLP is trained on input-output pairs and models the relationship between the two by adjusting weights and biases of the model to minimize errors. The adjustments are made via back propagation relative to the error (mean squared error, MSE, in this case).

Both the RF and MLP models are trained using current and recent cloud data from the ICI: 1) cloudiness of the ICI pixel for a chosen LOS; 2) cloud fraction for a small region around the LOS; and 3) cloud fraction of the entire ICI skydome. Transmission loss derived from the CL51 is also using in training. The models are trained using data from the period of August 2017 – June 2019, and inferenced over an independent period of July 2019 – June 2020. All data is preprocessed for quality assurance, and missing data is identified so as not to be included in the training of the model. Additionally, the data is normalized, so all predictors have approximately the same range of values. The data is separated into training and inferencing blocks based on the length of the database, the data lookback window, and the number of forecasts. The lookback window is defined by the period of time leading up to the current time and has been optimized at 60 minutes. Forecasts produced by the RF and MLP models predict whether or not the LOS will be cloudy at any time within a forecast range. For example, the 60 minute forecast does not predict whether it will be cloudy at exactly 60 minutes in the future, but rather if any time is cloudy from the forecast initialization time out to 60 minutes.

Both the RF (Fig. 1, top right) and the MLP (Fig. 1, bottom left) outperform the persistence cloud forecast (Fig. 1, top left). Because the site is clear so often, the persistence forecast has a strong clear bias, especially in longer forecasts. The performance of the MLP is very slightly better than the RF at the longer forecast lengths of 30 and 60 minutes.

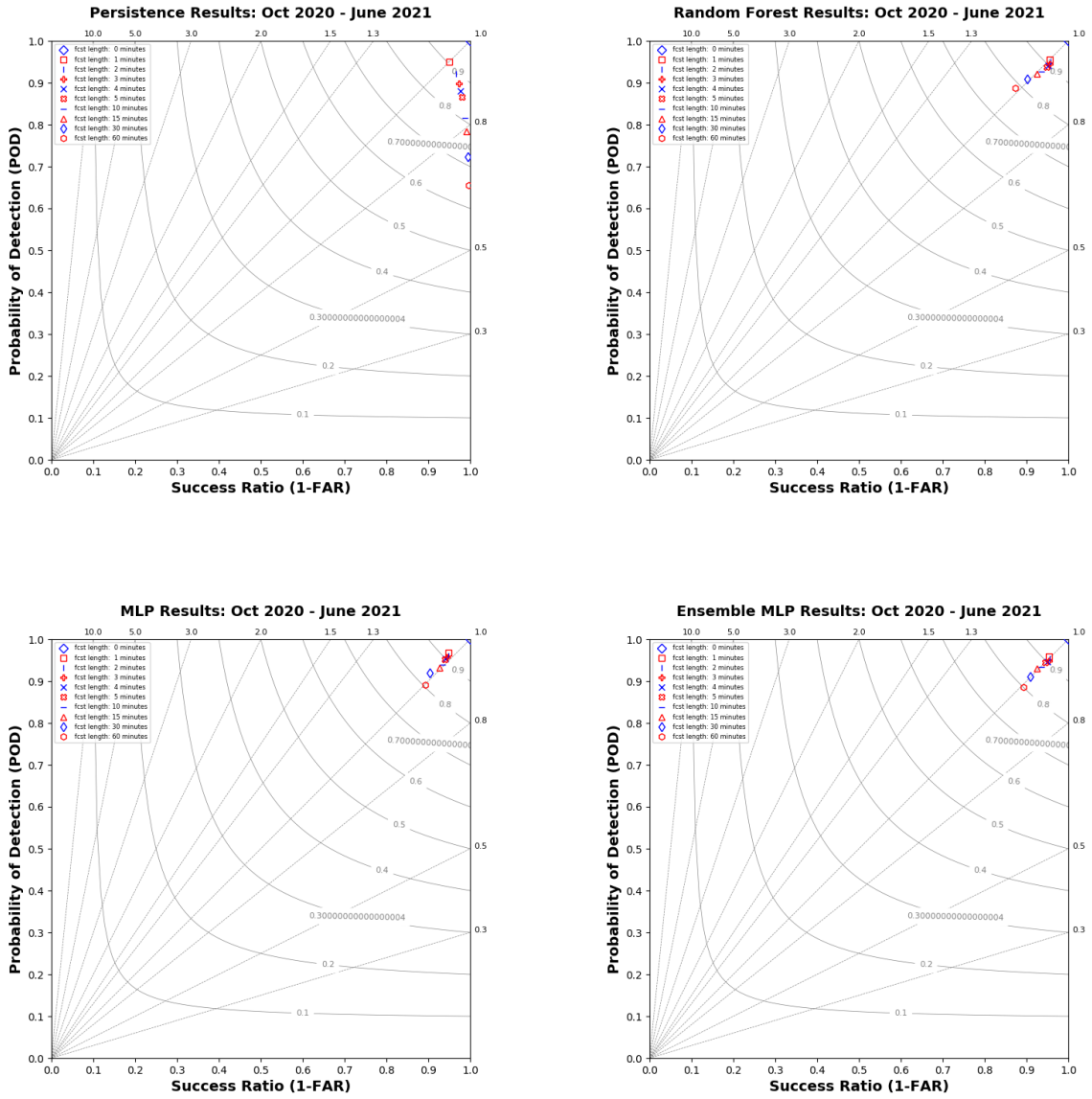

Fig. 1. Roebber plots showing the performance of persistence (top left) and the various ML models for forecast intervals of 0 minutes through 60 minutes. Markers in the top right corner indicate a perfect forecast. Markers along the diagonal indicate no bias. A clear bias is below and to the right of the diagonal and forecasts above and to the left of the diagonal indicate cloudy bias. The optimal forecast is the ensemble shown in the bottom right panel.

Since the performance of the RF and MLP are similar, an ensemble of these forecasts was developed to determine whether an additional application of MLP using a combination of the RF and initial MLP cloud predictions generates a more accurate forecast. This ensemble MLP is an enhanced forecast using the best qualities of each input forecast. Similar to the MLP and RF models, the ensemble uses information from the most recent cloud predictions as well as the predictions produced by the MLP and RF models during the previous 60 minutes. The ensemble model also incorporates cloud transmission loss estimates derived from CL51 backscatter values. Indeed, the ensemble acts to reduce the small cloudy bias of the MLP at the shortest forecast lengths, while maintaining the CSI and low bias at the longer forecast lengths.

These three models, the RF, MLP, and ensemble MLP, have been running operationally for the summit of Haleakala since October, 2020. Predictions are generated less than 60 seconds after an ICI image is taken. The cloud probabilities from the ensemble MLP are used to produce ten predictions out to 60 minutes. The clear/cloud probability is thresholded at 0.45 for 0-10 minute predictions, 0.35 for 30 minute predictions, and 0.30 for 60 minute predictions. These variable thresholds are chosen to minimize bias while maximizing the CSI. Probabilities under 0.10 are considered high confident clear (green), probabilities over 0.90 are considered high confident cloudy (red), and probabilities in between are more uncertain (yellow). The model most often has very high or very low cloud probabilities. From October, 2020 – June, 2021 fewer than 10% of 0-10 minute predictions are uncertain, and about 16% of 60 minute predictions are uncertain. The operational MLP ensemble model predictions are accurate more than 99% of the time for predictions 10 minutes and under, more than 97% for 15 minute predictions, more than 96% for 30 minute predictions, and more than 94% of the time for 60 minute predictions. Fig. 2 shows an example of the decision aid for an operational short-range cloud prediction with the corresponding ICI image.

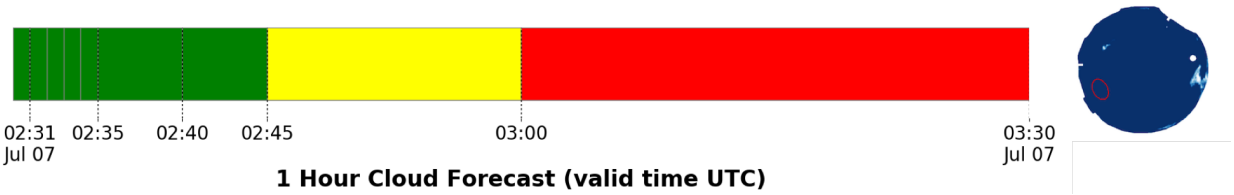

Fig. 2. A real-time ensemble MLP prediction from July 2021. The 1-15 minute prediction is for a clear LOS, the 30 minute prediction is uncertain, and the 60 minute prediction is cloudy. The ICI image shows the clouds on the right that are predicted to block the LOS within 60 minutes.

3. UNET CLOUD ANALYSIS

Predictions of clouds for the mid-range timescale of 0 to 2 hours are based on cloud analyses derived from satellite data from the NOAA GOES-17 Advanced Baseline Imager (ABI). Analysis of the GOES-17 Level 2 Binary Cloud Mask (BCM) product indicates systematic over-detection of clouds particularly over high altitudes at night. For example, erroneous star-shaped cloud patterns are visible over Haleakala, Mauna Kea and Mauna Loa, HI (Fig. 3). The errors in this product are likely due to the instrument's well-known cooling system issue where inadequate cooling causes infrared channel saturation for several hours at night before and after the seasonal equinoxes. Consequently, the BCM product is deemed unsuitable for accurate predictive capabilities.

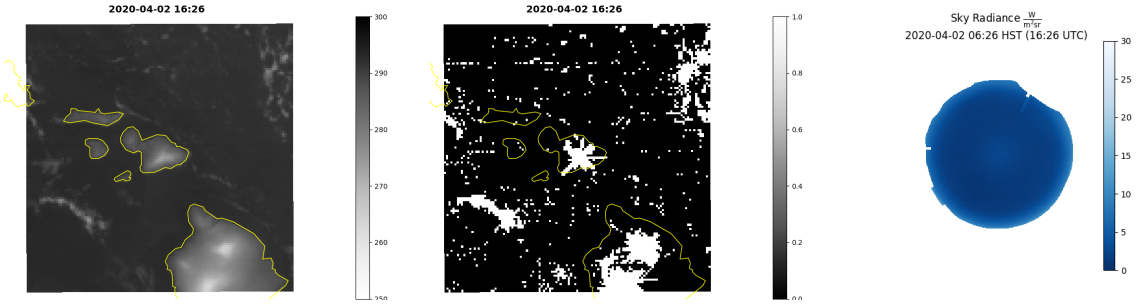

Fig. 3. GOES-17 Channel 13 (left), GOES-17 BCM product (middle) and ICI (right) over Haleakala. False detection of clouds at high elevations where IR temperatures are cool shown in the BCM image. Independent instrument measurement from ICI on the summit of Haleakala indicates clear sky.

Fortunately, the impact of the ABI's cooling system is not uniform across all of the infrared channels and not all channels are necessary to produce accurate cloud analyses. Taking advantage of this, a new cloud analysis based on the two ABI infrared channels least impacted by inadequate cooling at night and a visible channel was developed. The U-Net convolutional neural network is known for its speed and capability for pixel-based semantic segmentation and its success with limited data, made possible by applying neural computer vision rather than a

pixel-radiometric approach [7]. It is well suited for cloud detection as it can readily be architected to label each pixel with a probability of cloudiness.

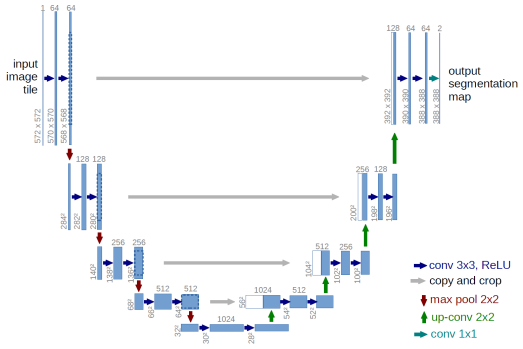

Fig. 4. The U-Net architecture has two paths, the encoder path that contains convolutional and max pooling layers which determines what information is in an image, and a decoder path that determines where image features are and is used to assemble a precise localized output.

The cloud analyses training data consists of 5 predictors including: the ABI's visible channel 2 (0.64µm), shortwave infrared channel 7 (3.9µm), and longwave infrared channel 13 (10.3µm); the United States Geological Survey's (USGS's) GTOPO30 terrain height; and the solar zenith angle. The data is remapped to a fixed 1 km, Cartesian latitude/longitude grid of size 256x256, centered on Haleakala, HI, to account for the different spatial resolutions between the visible and infrared channels. The cloud analyses labeled data consists of binary cloud masks produced as part of a 25-year climatology derived from Geostationary Operational Environmental Satellite (GOES) West satellites, GOES-9, 11, 15, from 1995-2020 [5]. The labeled cloud masks are 1 km spatial and 15 minute time resolution. The training period of record is 9 months spanning from June 2019 through February 2020, when GOES-15 and GOES-17 overlapped coverage over HI. Data is quality controlled and screened through several tests to identify and remove scanlines and noisy data, and then normalized. There are a total of 13,000 training and 6,000 validation samples yielding 1.9 million trainable parameters. The model is inferenced with over 2,000 samples.

The U-Net's internal convolutions use a leaky ReLU (Rectified Linear Unit) activation function with batch normalization and L2 regularization=0.001 to converge to a solution. The Adam optimizer is used and the final activation function is a sigmoid function with the loss function defined as mean squared error. The initial learning rate is $1e^{-4}$ with step-down on plateau to $1e^{-5}$ enabled.

Analysis of the UCA has proven to be a huge success, producing cloud analyses equal to the GOES-17 BCM during the day and consistently better than the BCM at night, and, at times, even better than the labeled dataset. The 1-minute sky radiance measurements from the Infrared Cloud Imager (ICI) installed on the summit of Haleakala are used to provide an independent assessment of clouds at night, when the GOES satellite is known to have accuracy issues. The ICI dataset is also used as an additional evaluation source for the UCA. Fig. 5 shows the UCA's performance statistics for the 256x256 grid and the pixel that covers Haleakala as compared to the historical cloud climatology. The probability of detection of the domain is 84% with a CSI of 88% and no bias. The site performance at the summit of Haleakala is slightly lower with a cloudy bias due to reduced cloud detection over the summit in the labeled dataset.

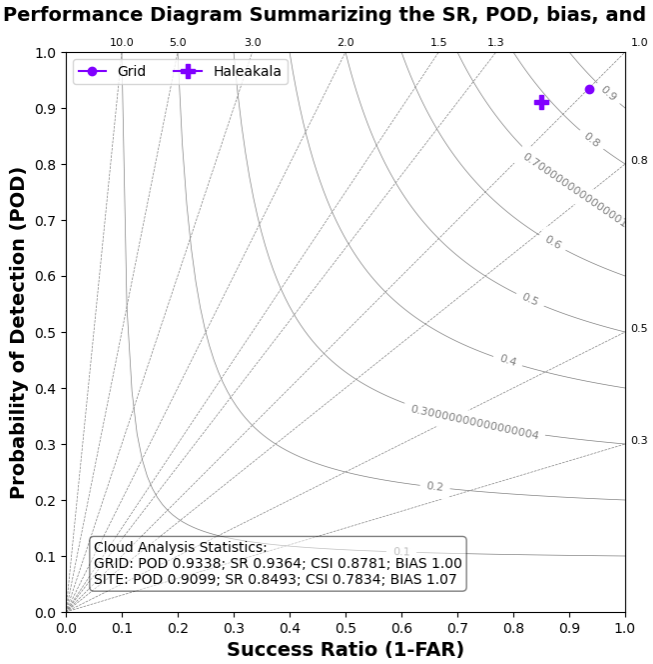

Fig. 5. Roebber statistics of the UCA 256x256 domain and the pixel that covers Haleakala as compared to the labeled cloud data.

A nighttime example of the UCA derived from the GOES-17 Channel 13 data and Channel 7 data as compared to the GOES-17 BCM dataset is shown in Fig. *6*. The UCA's accurate clear sky analysis over Haleakala is confirmed by the ICI measurement at the same time. The UCA also has a general lack of noise as compared to the BCM data at this same timestep. During the day, the UCA takes advantage of the availability of the GOES-17 visible channel 2. In the example shown in Fig. *7*, the cloud cover over Maui is accurately identified by the UCA and the summit is correctly predicted clear, which is confirmed by the independent measurement of sky radiance from the ICI. While the GOES-17 BCM daytime performance is generally good, this example reflects an over detection of clouds.

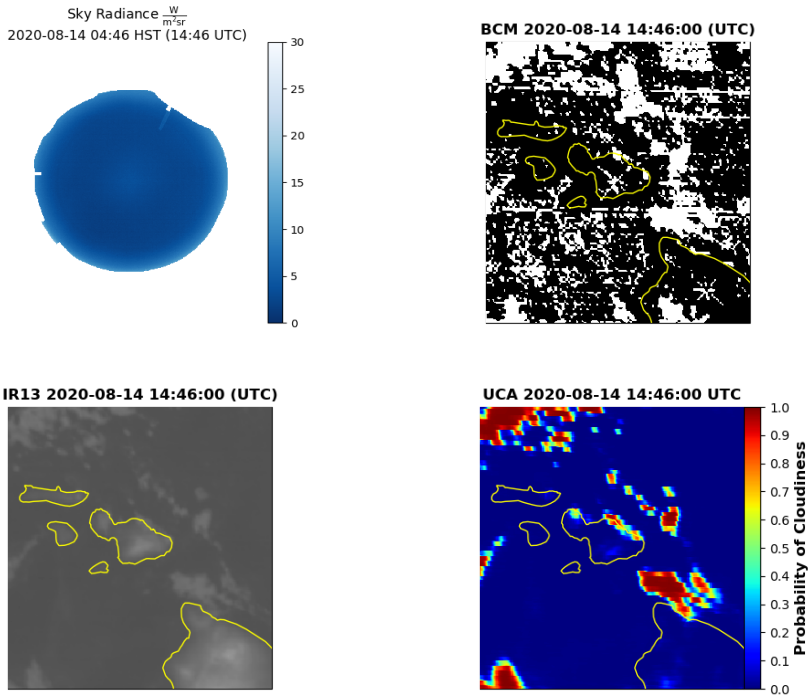

Fig. 6. Nighttime example of the UCA, valid at August 14, 2020 14:46 UTC. Top row: ICI, GOES-17 BCM product. Bottom row: GOES-17 infrared channel 13, UCA analysis

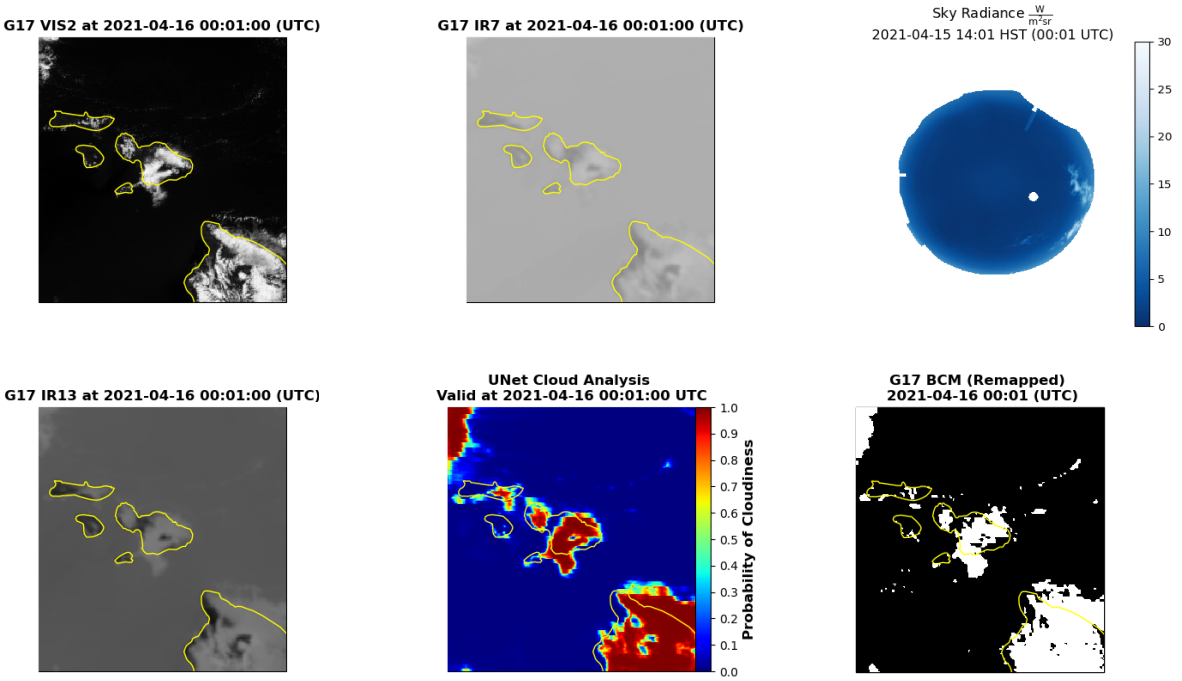

Fig. 7 Daytime UCA example, valid at April 16, 2021 at 00:01 UTC. Top row: GOES-17 visible channel 2, GOES-17 infrared channel 7, ICI sky radiance. Bottom row: GOES-17 infrared channel 13, UCA, and GOES-17 BCM

4. UNET CLOUD FORECAST (UCFS)

After generating the UCA for a 13-month period, spanning from March 2020 through March 2021, a cloud forecast based on this data and the U-Net is developed. It is subsequently referred to as the U-Net Cloud Forecasting System (UCFS). The UCFS has 15 minute time resolution and currently predicts clouds out to 2 hours for a 256x256 domain. The input to the U-Net model consists of the UCA for the current image time and the previous seven image timesteps at 15 minute resolution. Since the data is a probability of cloud, no additional normalization is required. The output is a prediction of the probability of cloud at each pixel for the next two hours in the future.

The U-Net architecture is largely consistent with the architecture used for the UCA utilizing the same number of layers, activation functions, and loss function. The batch size, the number of base filters, and the learning rate are increased to support the larger input and output dimensions of the UCFS. The training and validation data consist of the 13 months of UCA data, yielding 51,000 training and 22,000 validation samples. Inferencing is performed on independent months beginning in May, 2021 with ~7,000 samples in per month.

The project's metric for success is chosen to be whether or not the CFS performs better than a persistence forecast, specifically at the Haleakala summit. The Roebber statistics in Fig. 8 (left) show that the CFS beats persistence at all forecast lengths from 15 minutes to 2 hours from analysis time with little bias, and that performance increases with forecast length. Fig. 8 (right) represents the UCFS calibration accuracy and indicates that the model performs well, particularly under fully clear or cloudy skies.

The UCA and the UCFS are both running operationally as of April 2021. The NOAA GOES-17 data is ingested every five minutes, remapped to a 1 km grid, and then inferenced using the U-Net UCA model to produce a real-time cloud analysis. This cloud analysis, along with the previous seven cloud analyses at 15 minute resolution, are used as input to the U-Net UCFS model, which then produces a real-time prediction out to 2 hours. All UCA and UCFS data are archived in NetCDF files in order to expand the models' training datasets going forward. At the end of each month, the UCA and UCFS performance metrics are assessed.

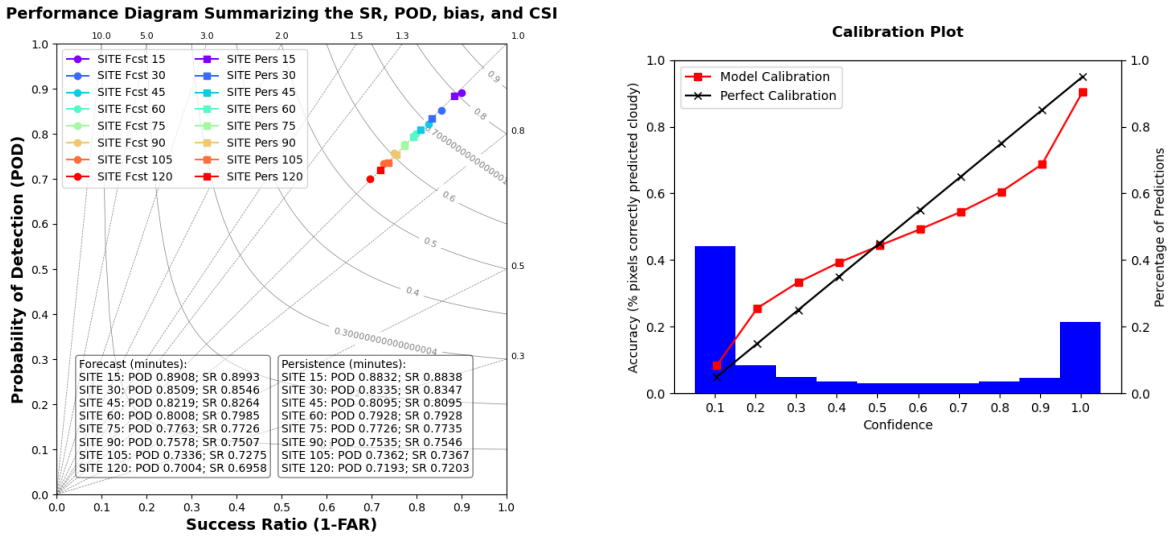

Fig. 8. Statistics over Haleakala indicate UCFS out-performs persistence at most forecast lengths. Reliability diagram (right) depicts the predicted cloud probability (red) versus the observed relative frequency of clouds (blue). Ideally, these values should be the same, with the points falling on the 1:1 line. The UCFS model has highest accuracy when 100% clear or cloudy scenes are predicted, which occurs the majority of the time (right).

5. LONG TERM CLOUD FORECAST

As with the short- and mid-range predictions, long-range cloud predictions can be made using various methods including persistence and climatology. Although these simple methods may work reasonably well for some locations with relatively static weather patterns, a better approach is to use Numerical Weather Prediction (NWP). NWP models use sophisticated dynamical models to predict meteorological variables within a regional or global grid, and are used by weather agencies to make daily weather forecasts. However, these global and regional weather forecasts are generally too coarse in spatial resolution to provide enough accuracy at the highly local scale that is required for space-based laser and surveillance applications. This work shows the benefits of applying AI techniques to the predictions of a high-resolution, regional NWP model.

The potential of Machine Learning to improve NWP cloud predictions has been shown in recent years [1, 2]. For this work, the Weather Research and Forecasting (WRF) model version 3.6 (Skamarock et al. 2008) is used to produce high resolution forecasts of meteorological parameters, including clouds, over the summit of Haleakala. The WRF model configuration 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. 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.

Despite its high resolution, the WRF forecasts have been shown to under-predict clouds (have a clear bias) at the summit of Haleakala [1]. In this work, two AI techniques are used to correct the under-prediction of clouds by WRF: the combination of an autoencoder and a U-Net on the entire 1/3 km WRF domain; and an MLP trained only on the Haleakala summit. In each case, the labeled data used in training and validation is a cloud analysis, derived from the GOES-15 satellite, consisting of a two-dimensional binary array of cloudy or clear decisions at 1 km spatial and 1 hour temporal resolution.

The predictors for the U-Net model are generated from the WRF 48 hour forecasts for the 1/3 km inner domain for the period of August 2017 – 2018. The WRF output files are preprocessed to extract atmospheric parameters, including the forecasts of relative humidity (RH), temperature, winds, cloud fraction and several other derived parameters, on 24 pressure levels. The 24 vertical levels and 8 atmospheric state variables (192 combinations) are consolidated into groups of one to nine variables using an autoencoder. This consolidation is analogous to using principle components to retain the greatest share of information about the multi-variate vertical atmospheric states possible in a few coded variables. However, autoencoders can be more efficient at this task than conventional principle components. In this case, the autoencoder is a symmetric bottle-neck multi-layer perceptron, which is trained to take the 192 dimensional vertical state, and then recreate it with minimal error after passing through the bottleneck. Once trained, the interior activation levels of the bottleneck-layer are the encoded values. Unlike principle component analysis, these values are unordered, and hence, to have a best n-codes representation, one needs train a specific autoencoder with a bottleneck layer of size n. It is hypothesized that using a n-codes autoencoded representation of the atmospheric state would have as much or more utility in forecasting future state as any select of n specific input variable-level combinations, and most likely, any simple set of n combinations of variables. As with any deep learning process, careful attention was paid to the hyper-parameterization of the autoencoder itself, to the selection of training, validation, and testing data sets, and to the computational training strategy used. These codes are the set of predictors that are used in training the WRF-based U-Net cloud forecast model, and are stored in NetCDF files along with the labeled cloud analysis data derived from GOES-15.

The U-Net model uses the Adam optimizer, an initial learning rate of 0.0001 with the option of reduce learning rate on plateau enabled, batch size of 16, three U-Net layers, 16 filters, and the MSE loss function. For the training, the data is divided into 70% used for training and 30% for validation. A separate dataset comprised of 9 months of data from March, 2019 – December, 2019 is then inferenced with the trained U-Net model, from which the accuracy statistics shown here are computed. The left and middle panels in Fig. 9 show that the U-Net generally captures the average cloud cover over Haleakala. The Roebber statistics of the summit of Haleakala are displayed in the right panel of Fig. 9. This shows the clear bias of the native WRF cloud forecast at the summit (green circle), and its relatively low CSI of 0.3. The U-Net does improve upon the WRF cloud forecast, producing no bias at the summit,

but the U-Net CSI is only marginally better than WRF. Both the WRF and U-Net predictions significantly outperform a simple persistence forecast.

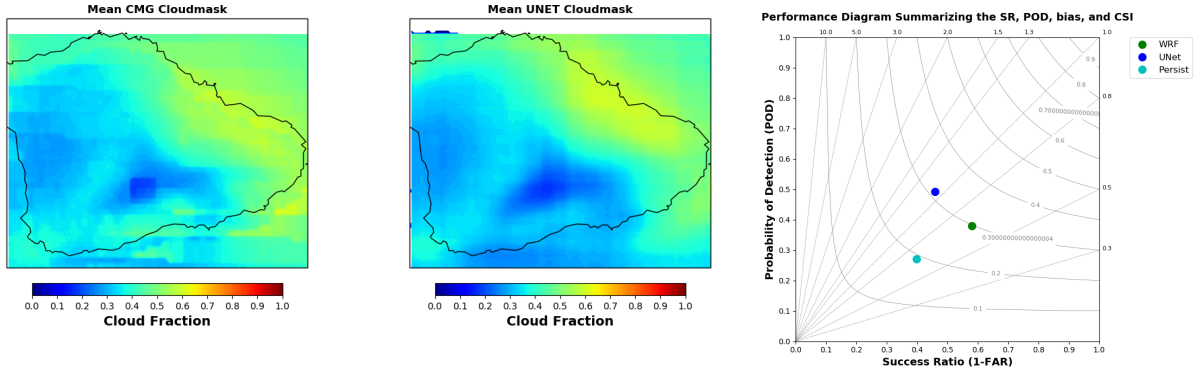

Fig. 9. The mean satellite-derived (left) and WRF-based U-Net (middle) cloud fractions. Roebber statistics of the WRF (green), U-Net (blue), and persistence (cyan) cloud forecasts for the Haleakala summit.

Previous work has shown that a simple Random Forest technique can significantly improve upon the native WRF cloud fraction by focusing the training on a small region within the WRF domain. Building upon this work, an MLP is used to enhance the WRF cloud forecast by focusing the training on the WRF forecast data local to the Haleakala summit, ignoring the rest of the WRF domain. Many of the same atmospheric parameters from the WRF forecasts as are used in the U-Net training are used to train the MLP. These include surface and three-dimensional parameters from a small area around and above the Haleakala summit extracted from the WRF forecast files to be used as predictors for the MLP training. The MLP is trained using the same satellite-derived cloud analysis as the labeled data for the period August 2017 – February 2019, and inferenced over March 2019 – December 2019.

The results show a significant improvement over both the WRF and U-Net cloud predictions for the Haleakala summit (Fig. 10). The CSI improves to 0.42, while the cloud predictions remain unbiased (Fig. 10, left panel). As previously discussed, the CSI is designed to provide an indication of how well events are predicted. In this case, the occurrence of clouds is the event, and correct predictions of the absence of clouds are not considered in the calculation of CSI. Therefore, the overall accuracy of the MLP is also compared to the WRF accuracy in Fig. 10 (right panel) to supplement the Roebber statistics. The MLP-enhanced cloud predictions are consistently 5-10% better than the native WRF cloud forecast. These two figures show the MLP not only corrects the clear bias in the native WRF cloud forecasts, but also does significantly better at predicting both the occurrence and absence of clouds.

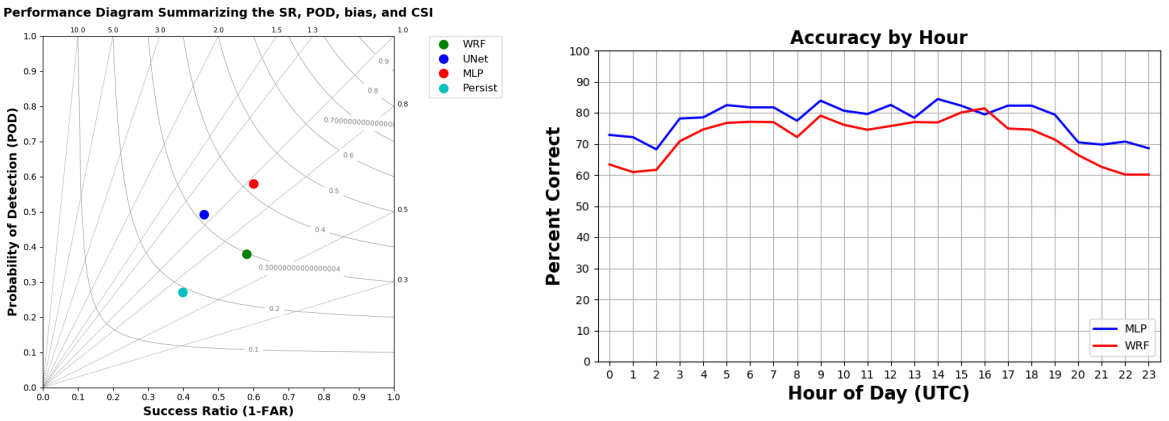

Fig. 10. Performance statistics for the WRF, U-Net, MLP, and Persistence cloud predictions (left) and the overall accuracy of the MLP and WRF cloud predictions (right). The MLP outperforms both WRF and U-Net for the summit of Haleakala.

Since Haleakala is a unique, isolated tropical summit not at all typical of many locations, the U-Net and MLP techniques were also applied to a mid-latitude location in Southern California to determine how they compare to the Haleakala results. Another high-resolution WRF domain is run daily over the southwestern United States, also producing 48 hour forecasts. The data from these WRF forecasts are prepared for the U-Net and MLP training with the same method used for the Haleakala region, except it is applied to the WRF domain with 3 km resolution instead of the innermost WRF domain. The U-Net is trained over the entire domain, and the MLP is trained for a single location.

As might be expected for a mid-latitude location impacted by larger weather systems, the native WRF cloud forecasts perform better than for the isolated Haleakala summit. The WRF cloud forecast CSI is 0.46, and has a small clear bias (Fig. 11). The U-Net, trained over the entire WRF domain, produces a CSI very similar to the WRF cloud forecast CSI. As with Haleakala, the MLP, trained for a single location, outperforms both the WRF and U-Net for that location, generating a CSI of 0.59.

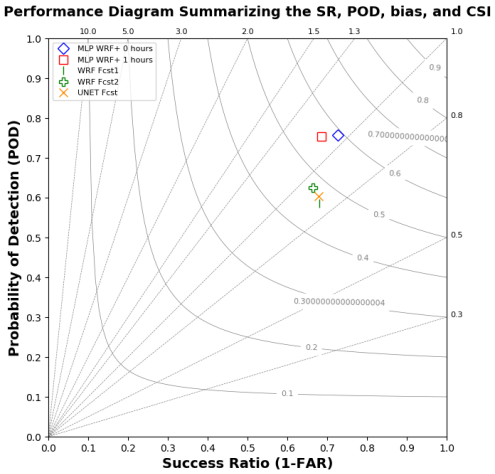

Fig. 11. Performance statistics for the WRF, U-Net, and MLP cloud predictions for the Southern California location. Both the WRF (green symbols) and the U-Net cloud predictions have CSI (orange ×) values near 0.46 with a small clear bias. The MLP outperforms WRF and the U-Net with a CSI of 0.59 (red and blue symbols).

6. DISCUSSION

The accurate characterization and prediction of atmospherics, and clouds in particular, is vital to space-based laser and surveillance applications. Mission decisions are made at different time scales, from short-range, minute-to-minute link handover decisions, to longer-range schedule and maintenance planning. This work demonstrates that AI-powered cloud predictions consistently and significantly outperform both persistence and NWP cloud predictions. Once the data is staged, the model training is accomplished in hours and days (not weeks and months) by taking advantage of accelerated processing on GPUs.

Using several sources of data, distinct DL models were developed to improve upon existing cloud predictions at three different time scales. An ensemble MLP is used to predict clouds out to one hour using ground-based ICI and CL51 cloud data. A mid-range cloud prediction is generated by applying a U-Net convolutional neural network (UCFS) to satellite-derived cloud data, which in turn is generated by the UCA, a U-Net model that improves upon the existing GOES cloud analysis. Finally, long-range cloud predictions are generated by an MLP to enhance the daily WRF forecasts for the summit. In each case, significant effort was spent pre-processing the labeled and validation datasets, performing quality assurance on the data, and optimizing the hyperparameters.

It is important to note that the accuracy of each technique is limited in some amount by the quality of the labeled datasets. Each has its limitations; for example, GOES satellite products at night and ICI cloud analyses are influenced by atmospheric water vapor loading. To this end, we are currently leveraging our experience to develop an improved ICI cloud analysis using a U-Net, and investigating ways to improve the UCA. These new products and larger datasets should help to continue to improve the fidelity of the multi-scale atmospheric prediction system.

7. REFERENCES

[1] Russell, A.M., R.J. Alliss, and B.D. Felton, "A Deep Machine Learning Algorithm to Optimize the Forecast of Atmospherics", Advanced Maui Optical and Space Surveillance Technologies Conference, 2017.

[2] Felton, B.D., "Deep Learning to Improve Numerical Weather Prediction Cloud Forecasts", in *19th Conference on Artificial Intelligence for Environmental Science*, American Meteorological Soc., January 2020.

[3] Alliss, R.J., M. E. Loftus, D. Apling, and J. Lefever, "The Development of Cloud Retrieval Algorithms Applied to GOES Digital Data," in *10th Conference on Satellite Meteorology and Oceanography*, pp. 330–333, American Meteorological Soc., January 2000.

[4] Wojcik, G., R.J., Alliss and M.E., Craddock, "Deep Space to Ground Laser Communications in a Cloudy World", *IEEE Photonics*, Aug 2005.

[5] Alliss, R.J., H.S. Kiley, M.E. Craddock, and B.D. Felton, "Atmospheric Characterization of the Space Environment: Unique Observations from Haleakala", Advanced Maui Optical and Space Surveillance Technologies Conference, 2019.

[6] Roebber, P.J., Visualizing Multiple Measure of Forecast Quality, *Weather and Forecasting*, Vol. 24, 601–608, 2009.

[7] Ronneberger, Olaf; Fischer, Philipp; Brox, Thomas (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation", arXiv:1505.04597.