16D.3

CENTER FIXING TROPICAL DEPRESSIONS AND TROPICAL STORMS USING MACHINE LEARNING-NIGHTTIME VISIBLE IMAGERY

Nathan Stanford*
Air Force Institute of Technology

1. ML-NVI

The Machine Learning-Nighttime Visible Imagery (NVI) algorithm is a Fully Connected Neural Network consisting of ten input features, three hidden layers (two with eight nodes and one with four), and a single output layer node (Pasillas 2023). The ten input features are derived from four infrared (IR) wavelengths and their brightness temperature differences (BTD)s (Pasillas 2023, 14). Although originally trained and tested on Visible Infrared Imaging Radiometer Suite (VIIRS) IR channels, the wavelengths selected closely align with geostationary satellite channels, which allowed for validation using Geostationary Operational Environmental Satellite (GOES) Advanced Baseline Imager (ABI) (Pasillas 2023). The three hidden layers use Rectified Linear Unit (ReLU) activation, which does not saturate for positive values, an Adam optimizer of .001 (the default value), and a mean square error loss function (Geron 2022; Pasillas 2023). The single output layer, which uses a Sigmoid activation function, computes a pixel-by-pixel synthetic lunar reflectance value representing lunar reflectance values expected during a full moon even in the absence of moonlight (Pasillas 2023, 16).

NVI uses the VIIRS Day/Night Band (DNB) Miller-Turner Lunar Reflectance values as truth data (Miller and Turner 2009; Pasillas 2023). It requires no scaling because it is based on a lunar spectral irradiance data set, which makes it possible to calculate a quantitative value ranging 0-1 (Miller and Turner 2009; Pasillas 2023). Pasillas (2023) developed two NVI models, one using VIIRS Bands 13 (4.05μm), 14 (8.55μm), 15 (10.763μm), and 16 (12.01μm), and the other using ABI Bands 7 (3.9μm), 11 (8.4μm), 13 (10.3μm), and 15 (12.3μm). The VIIRS model demonstrated a root mean square error (RMSE) of 12.05 (lunar reflectance), while the ABI model exhibited RMSE of 16.2 when compared to true lunar reflectance values at full moon (Pasillas 2023, 45).

The principal function of NVI in tropical cyclone (TC) forecasting is highlighting upper-level cirrus clouds increasing the clarity of features associated with the low-level circulation center. **Figure 1** shows NVI imagery in comparison to shortwave infrared (SWIR) (3.9μm) and longwave infrared (LWIR) (11.2μm). The white box in image 1 emphasizes how NVI clearly distinguishes upper level cloud features with distinct reflectivity differences for thick upper level clouds (white) and thin cirrus (grey). In contrast, the SWIR (2) and LWIR (3) images fail to distinguish the upper level features. The white box in image 4 demonstrates NVI's ability to delineate cloud phase, effectively “seeing through" the upper level cirrus clouds providing a clear visual of the TC eye; this eye clarity is not observed in SWIR (5) or LWIR (6).

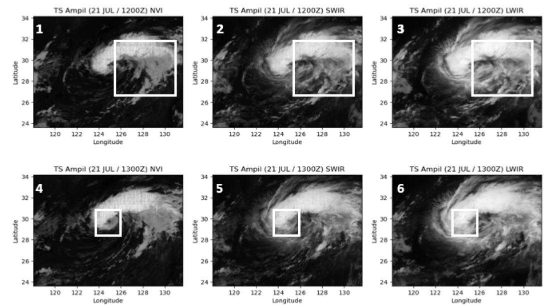

FIG. 1. Time series of NVI (1,4), SWIR (2,5), and LWIR (3,6) imagery for TS Ampil on 21 July 2018 at 1200Z and 1300Z.

2. ARCHER-2

ARCHER-2 utilizes TC scenes from microwave (specifically 37 and 85-92 GHz), SWIR (3.9μm), LWIR (11.2μm), visible (.64μm), and Advanced Scatterometer (5.25GHz) to determine an objective center fix. It does this through pattern recognition based on an objective analysis of a combined spiral and ring score, as well as a minor distance penalty (Wimmers and Velden 2016, 198). The spiral score assesses the center of the TC by computing a gridded vector field to measure the relative alignment of the brightness temperature gradient of the image and a spiral (or curved banding) pattern (Wimmers and Velden 2016, 198). The ring score evaluates the relative alignment of the brightness temperature gradient and an eye-like ring pattern by calculating the dot product of radially oriented unit vectors and the image gradient (Wimmers and Velden 2016, 198). Finally, ARCHER-2 calculates a minor distance penalty, which varies as the distance squared from a given initial center position guess, to reduce the probability of center fixing on peripheral features (Wimmers and Velden 2016, 198). For operational analysis, the initial center position guess is based on a short-term forecast from an operations center interpolated to the time of the satellite image (Wimmers and Velden 2016, 198). In retrospective studies, such as this one, an interpolation from a best track reanalysis is advised (Wimmers and Velden 2016, 198). The weighted average of these three metrics - spiral score, ring score, and distance penalty - constitute the combined score; the position of the maximum combined score is the center fix (Wimmers and Velden 2016, 199). **Figure 2** provides an example of the ARCHER-2 spiral, ring, and combined score contours. The white plus sign is the initial center

**Corresponding author address:* Nathan Stanford, Dept. of Engineering Physics, WPAFB, OH 45409; E-mail: nathan.stanford.1@us.af.mil.

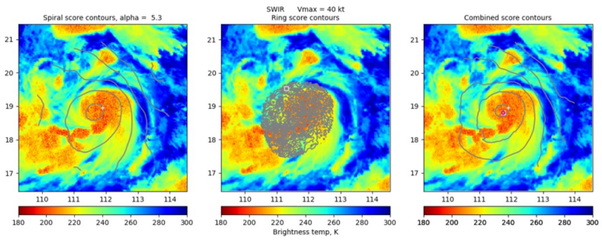

FIG. 2. ARCHER-2 spiral (left), ring (middle), and combined score (right) contours for TS SonTinh on 17 July 2018 at 1600Z.

position guess; the magenta circle in the combined score image (right) is the center fix location.

For each forecast period, ARCHER-2 assesses a center fix position for each imagery type, then selects the sensor with the greatest center fix certainty, referred to as the source sensor, for use as the center latitude and longitude in TC retrieval algorithms. Center fix certainty is a function of the expected error, which follows the gamma distribution provided in **Equation 1**:

$$PDF(x) = \alpha^2 x e^{-\alpha x} \tag{1}$$

where α characterizes the distribution of the expected error (x) (Wimmers and Velden 2016, 202). In addition to the α parameter, Wimmers and Velden (2016) assessed that the best indication of accuracy is the spacing of the combined score contours surrounding the center fix location; therefore, the tighter the contour spacing, the lower the expected error for each sensor (Wimmers and Velden 2016, 202). This led to the development of the Confidence Score (η), which assesses the difference between the maximum combined score (or center fix) and the highest score $\geq .75^o$ latitude and longitude from the center fix location (Wimmers and Velden 2016, 202). Alpha and confidence score were then used to calibrate the expected error of a center fix, as shown in **Equation 2**:

$$\alpha(\eta) = m_{sensor}\eta + b_{sensor} \tag{2}$$

where m_{sensor} and b_{sensor} are the sensor-specific slope and offset, respectively. Ultimately, it is the highest alpha score that determines the source sensor for a forecast period. Sensor-specific fit parameters between alpha and confidence score are provided in **Table 1** (Wimmers and Velden 2016, 204).

ARCHER-2 Geostationary Sensor Parametric Fit				
Sensor	m_{lo}	b_{lo}	m_{hi}	b_{hi}
DNB/Visible	14.44	-0.83	14.64	3.45
SWIR	8.68	-0.37	14.2	-0.24
LWIR	9.89	-2.07	9.26	1.95

TABLE 1. Sensor specific slope and offset values used by ARCHER-2 to relate confidence score and alpha. The "lo" subscript applies to storms with maximum velocity below 65 knots; "hi" applies to storms greater than 85 knots. For maximum wind velocities between 65 and 85 knots, alpha is the weighted average for both "lo" and "hi".

Because a parametric fit for NVI was not calibrated during this study, NVI lunar reflectance values were linearly scaled to SWIR brightness temperature (BT)s ranging from 130K-400K, then histogram matched to the SWIR imagery for the same analysis period (reference **Figure 3**). From the histogram matched NVI imagery, ARCHER-2 calculated spiral, ring, and combined scores, as well as many other statistics, using the SWIR parametric fit.

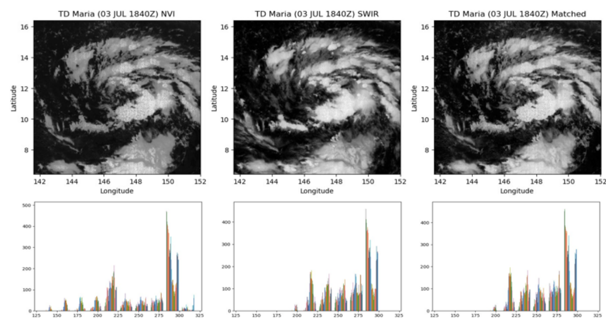

FIG. 3. Example of NVI (left) histogram matched to SWIR (middle). The matched image (right) exhibits similar image contrast as SWIR (reference image histograms in second row), while highlighting low-level stratus and upper level cirrus similar to the NVI image.

3. ARCHER-2 STATISTICS

Statistics for the study consist of raw output from ARCHER-2: Alpha, Rad50, Rad95, and Confidence Score. Additionally, two derived statistics were computed: Usage and Error. Alpha characterizes the distribution of the expected error and determines the source sensor. "Rad50" and "Rad95" are the 50% and 95% confidence radii measured in degrees latitude/longitude surrounding the center of the storm, respectively. Smaller "Rad50" and "Rad95" values indicate increased confidence in the

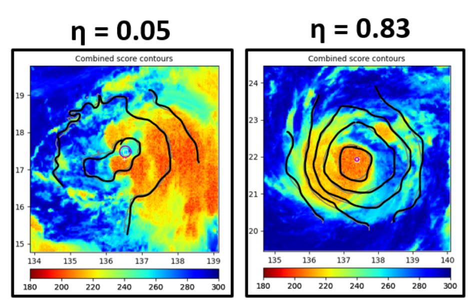

FIG. 4. Combined score contours for low (left) and high (right) confidence center fix analyses.

Overall Results			
Category	NVI	SWIR	LWIR
Confidence Score	0.22	0.27	0.26
Alpha	1.68	2.05	1.25
Rad50 (^{o})	1.74	1.50	2.42
Rad95 (^{o})	4.92	4.23	6.83
N	932	932	932
Usage (%)	32.6	54.1	13.3
Error (km)	39.8 (21-77)	60.4 (32-139)	99.7 (48-209)

TABLE 2. ARCHER-2 results for the ten storms assessed in this study. The upper section of the chart shows the mean values from ARCHER-2 raw output. The lower section shows the derived statistics: mean usage and median error with interquartile range.

center fix. Alpha, Rad50, and Rad95 share a fixed correspondence because they are derived from the same distribution curve; therefore, the higher the alpha score, the smaller the confidence radii. Confidence score (η) assesses the certainty of the ARCHER-2 center fix. The tighter the combined score contours, the higher the confidence score (reference **Figure 4**). Additionally, because alpha is a function of the confidence score (see **Equation 2**), the higher the confidence score, the higher the alpha is for a sensor.

For usage, ARCHER-2 selects the imagery type with the highest alpha score in an analysis period as the source sensor; therefore, the number of times an imagery type carries the highest alpha score out of the total analyses provides an overall usage percentage. The other derived statistic, error, derived via Haversine formula, is the measured distance between the ARCHER-2 center fix and the interpolated Best Track. All statistics, except error, provide mean values assessed at each analysis time, then averaged for the full data set in this study. Error follows a similar structure, except median values are provided due to the bimodal distribution of error across tropical depression (TD)- and TS-strength TCs.

4. RESULTS

The overall results for the objective study are provided in **Table 2**. Because of NVI's ability to assess cloud phase and differentiate upper and lower level cloud features, it exhibits lower median error than both SWIR (3.9μm) and LWIR (11.2μm) by 20.6 and 59.9 kilometers, respectively. SWIR demonstrates the optimal confidence score, alpha, Rad50, Rad95, and usage. This apparent contradiction between NVI's error and SWIR's confidence-related statistics likely arises from the decreased gradient stemming from NVI's cirrus handling, which causes cirrus clouds to appear smooth and grey (reference **Figure 1**). Additionally, by histogram matching NVI to SWIR imagery, then using ARCHER-2's SWIR parametric fit, the overall results for NVI's confidence score, alpha, Rad50, Rad95, and usage will contain uncertainty for the true output; that is, until a parametric fit for NVI is assessed.

To better measure uncertainty due to the parametric fit, NVI lunar reflectance values for the storms analyzed were also linearly scaled to brightness values (0-255), then converted into the ARCHER-2 variable, “pseudo-BT." This allowed ARCHER-2 to compute spiral, ring, and combined scores for NVI using the DNB parametric fit. **Equation 3** shows the pseudo-BT formula that ARCHER-2 uses to convert scaled brightness value into a pseudo-BT:

$$PT_B = 350 - .75(BV) \quad (3)$$

where BV is the brightness value. The comparison of ARCHER-2 raw output using NVI with DNB's parametric fit and NVI with SWIR's parametric fit is provided in **Table 3**. For NVI, mean alpha, Rad50, Rad95, and usage increased when using the DNB parametric fit; this is expected based on the sensor-specific slope and offset values for DNB (reference **Table 1**). The 0.3 kilometer increase in median error and 0.02 increase in confidence score between "DNB-fit" and "SWIR-fit" are attributed to scaling differences that arise when either converting NVI lunar reflectance to pseudo-BT for "DNB-fit" or histogram matching NVI to SWIR for "SWIR-fit". Ultimately, the results in **Table 3** provide a range of likely values, specifically for NVI's raw ARCHER-2 output (ie. alpha, Rad50, and Rad95).

4.1. 2016 ARCHER-2 Study Comparison

Comparing the results of this study with the 2016 ARCHER-2 validation, NVI's performance aligns closest with visible imagery for TD- and TS-strength storms (Wimmers and Velden 2016, 205). In Wimmers and Velden's 2016 study, visible, SWIR, and LWIR exhibited center fix median errors of 36, 46, and 56 kilometers, respectively (Wimmers and Velden 2016, 205). The 39.3 kilometer median error demonstrated by NVI highlights its capability to generate pseudo-visible nighttime imagery similar to standard visible (.64μm). Furthermore, any disparity in median error for SWIR and LWIR between Wimmers and Velden's study and this study likely stems from the TCs analyzed. Wimmers and Velden (2016) com-

Parametric Fit Uncertainty						
	DNB-fit			***SWIR-fit***		
	NVI	SWIR	LWIR	NVI	SWIR	LWIR
Confidence Score	0.24	0.27	0.26	0.22	0.27	0.26
Alpha	2.76	2.05	1.25	1.68	2.05	1.25
Rad50 (^{o})	1.39	1.50	2.42	1.74	1.50	2.42
Rad95 (^{o})	3.92	4.23	6.83	4.92	4.23	6.83
N	932	*	*	*	*	*
Usage	53.6	36.0	10.4	32.6	54.1	13.3
Median Error	40.1	60.4	99.7	39.8	60.4	99.7

TABLE 3. Raw and derived ARCHER-2 output using NVI with the DNB parametric fit versus the SWIR parametric fit.

bined TD and TS in a single category, so the sample size for each respective storm strength is unknown (Wimmers and Velden 2016, 205). From the error results though, it can be inferred that more TD-strength storms were examined in this study since the median error for SWIR and LWIR increased by 14.4 and 43.7 kilometers (Wimmers and Velden 2016, 205). Median error results for both the 2016 and current study are provided in **Table 4**.

2016 Study Comparison				
	NVI	SWIR	LWIR	
Imagery Type	NVI	Visible	SWIR	LWIR
2016 Study (km)	*	36	46	56
Current Study (km)	39.8	*	60.4	99.7

TABLE 4. Error results (in km) for Wimmers and Velden's 2016 study and the current study.

5. FUTURE WORK

The Joint Typhoon Warning Center (JTWC)'s research priorities called for an objective analysis of TD- and TS-strength storms in the Western Pacific (WPAC), specifically those that proved the most challenging to center fix. Future studies should extend this research to TCs of all strengths across all basins. Additionally, ARCHER-2 requires a parametric fit for NVI in order to use it operationally. In Wimmers and Velden's 2016 study, this amounted to five years of TC data across all storm intensities.

REFERENCES

Geron, A., 2022: *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.

Miller, S. D., and R. E. Turner, 2009: A dynamic lunar spectral irradiance data set for NPOESS/VIIRS Day/Night Band nighttime environmental applications. *IEEE Transactions on Geoscience and Remote Sensing*, **47 (7)**, 2316–2329.

Pasillas, C. M., 2023: Turning night into day: The creation, validation, and application of synthetic lunar reflectance values from the day/night band and infrared sensors for use with jpss viirs and goes abi. Ph.D. thesis, Colorado State University.

Wimmers, A. J., and C. S. Velden, 2016: Advancements in objective multisatellite tropical cyclone center fixing. *Journal of Applied Meteorology and Climatology*, **55**, 197–212.