
Tropical Cyclone Formation Probability (TCFP)

Algorithm Theoretical Basis Document

Compiled by the
**NOAA Center for Satellite Applications and Research
Tropical Cyclone Characteristics Science Team**



Version 1.0
January, 2024

TITLE: Tropical Cyclone Formation Probability ALGORITHM THEORETICAL BASIS
DOCUMENT VERSION 1.0

AUTHORS:

Chris Slocum (NOAA Center for Satellite Applications and Research)

TABLE OF CONTENTS

	<u>Page</u>
LIST OF TABLES.....	6
LIST OF FIGURES	7
1. INTRODUCTION	9
1.1. Product Overview	9
1.1.1. Product Description	9
1.1.2. Product Requirements	9
1.2. Satellite Instrument Description.....	10
2. ALGORITHM DESCRIPTION.....	11
2.1. Processing Outline	11
2.2. Algorithm Input	11
2.3. Theoretical Description.....	13
2.3.1. Physical Description	13
2.3.2. Mathematical Description.....	14
2.4. Algorithm Output.....	24
2.5. Performance Estimates	31
2.5.1. Test Data Description.....	31
2.5.2. Sensor Effects	31
2.5.3. Retrieval Errors	31
2.6. Practical Considerations.....	31
2.6.1. Numerical Computation Considerations.....	31
2.6.2. Programming and Procedural Considerations	32
2.6.3. Quality Assessment and Diagnostics	32
2.6.4. Exception Handling	32
2.7. Validation.....	32
3. ASSUMPTIONS AND LIMITATIONS	34
3.1. Performance Assumptions	34
3.2. Potential Improvements.....	34
4. REFERENCES	35

LIST OF TABLES

	<u>Page</u>
Table 1-1. NESDIS Product Baseline requirements for TCFP.	10
Table 2-1. TCFP guidance product algorithm input data with source, spatial and temporal coverage, and description.	12
Table 2-2. TCFP large-scale environment predictors.	17
Table 2-3. Mean Tropical North Atlantic Ocean atmospheric sounding of pressure, temperature, relative humidity, wind speed, and wind direction adapted from Table 1 in Dunion (2011). Published 2011 by the <i>American Meteorological Society</i>	18
Table 2-4. TCFP coefficient and averaging region and subregion names along with general longitudinal and latitudinal extent.	27

LIST OF FIGURES

	<u>Page</u>
Figure 2-1. A flow diagram of TCFP processing.....	11
Figure 2-2. A diagram depicting the algorithm input start times relative to algorithm run and processing create times. Color progresses from light to dark throughout the day.....	11
Figure 2-3. Graphic from Slocum et al. (2022) depicting GOES-16 10.3- μm longwave infrared observations at 0000 UTC 26 Aug 2017 showing Hurricane Harvey prior to landfall with the area-averaging regions for large-scale dynamics from 0 to 1000 km (blue circle), thermodynamics from 200 to 800 km (red annulus), and kinematics from 0 to 500 km (yellow circle). Published 2022 by the <i>American Meteorological Society</i>	15
Figure 2-4. The difference between azimuthal means for TCFP as a fix grid product in cyan (i.e., Eulerian) vs a storm position-following product like SHIPS in magenta (i.e., Lagrangian). The tan rings demonstrate two cases with averaging bands 100 km in radius wide around points to show overlap.....	15
Figure 2-5. A hodograph depicting the vertical wind profile from 1015 to 100 hPa (blue-shaded curve and black dots) and vectors for the 850–200-hPa deep-layer shear (red), 850–500-hPa shallow-layer shear (orange), and 850–200-hPa mass-weighted mean wind (yellow).	19
Figure 2-6. Vertical profiles of potential temperature (yellow curve), equivalent potential temperature (red curve), and saturated equivalent potential temperature (blue curve) based on the mean Tropical North Atlantic Ocean atmospheric sounding in Table 2-2. The black line indicates the value of the surface equivalent potential temperature (θ_e , SFC). The gray shaded area is the area used to calculate the vertical instability predictor. The gray dashed lines from bottom to top represent the lifted condensation level, level of free convection, and level of neutral buoyancy..	21
Figure 2-7. Vertical profiles of the (left) environmental temperature (yellow curve) and parcel virtual temperature (red curve) on a skew- T log- p diagram and (right) parcel vertical velocity as calculated by the entrained plume cloud model.	22
Figure 2-8. Global 0 to 24 h TCFP guidance product output tropical cyclogenesis probabilities on 1200 UTC 9 May 2023.	26
Figure 2-9. TCFP domain averaging large-scale regions.	26
Figure 2-10. TCFP domain averaging subregions.....	26
Figure 2-11. 0- to 500-km averaged cloud-cleared water vapor brightness temperature centered over the Atlantic Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.	28
Figure 2-12. 0- to 500-km averaged water vapor percent pixel below -40°C centered over the Atlantic Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.	28

Figure 2-13. 0- to 50-km averaged sea surface temperature centered over the Atlantic Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.	28
Figure 2-14. 200- to 800-km averaged environmental mean sea level pressure 0- to 50-km averaged sea surface temperature centered over the Western Pacific Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.	29
Figure 2-15. 0- to 1000-km averaged 850-hPa lower-tropospheric vorticity centered over the Western Pacific Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.	29
Figure 2-16. 0- to 1000-km averaged 850-hPa lower-tropospheric divergence centered over the Western Pacific Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.	29
Figure 2-17. 0- to 500-km generalized vertical wind shear centered over the Eastern and Central Pacific Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.	30
Figure 2-18. 200- to 800-km vertical instability centered over the Eastern and Central Pacific Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.	30
Figure 2-19. 0- to 500-km 850-hPa thermal wind temperature advection centered over the Indian Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.	30

1. INTRODUCTION

This document describes the Tropical Cyclone Formation Probability (TCFP) guidance product version 4.0, which provides hourly, machine-learning-generated probabilistic output for the likelihood of tropical cyclogenesis across the global oceans using a combination of geostationary satellite and numerical weather prediction information. NESDIS TCFP guidance product has roots in the Atlantic Tropical Cyclone Genesis Parameter (DeMaria et al. 2001). Schumacher et al. (2009) describe a 24-h product, later expanded to 48 hours that forms the basis for prior versions of TCFP. The fourth version of TCFP departs significantly from those methods, yet still provides probabilistic tropical cyclone genesis forecasts through 48h. NESDIS operationalized the fourth version in early 2024.

1.1. Product Overview

1.1.1. Product Description

The TCFP product generates the probability of genesis within 500 km of a point over two 24-h periods from the analysis time at grid points on a $1^\circ \times 1^\circ$ domain that spans from 0°E to 359°E and 45°S to 45°N . The probabilities originate from an equally-weighted consensus of calibrated machine learning models (i.e., linear discriminate analysis, logistic regression, random forest). The machine learning models use water vapor imagery from the global constellation of geostationary satellites to assess convective activity and large-scale subsidence, global sea surface temperature products for evaluating the available potential energy, and numerical weather prediction model-based information to determine the state of the large-scale environment by providing estimates of metrics such as vorticity, divergence, vertical wind shear, convective inhibition, and vertical instability. The product masks existing tropical cyclones using storm information from the NOAA National Weather Service National Hurricane Center in Miami, Florida, NOAA National Weather Service Central Pacific Hurricane Center in Honolulu, Hawaii, and Department of Defense Joint Typhoon Warning Center in Pearl Harbor, Hawaii. TCFP generates five types of output:

- 1) Probabilistic output in Network Common Data Form (NetCDF),
- 2) Probabilistic output in Keyhole Markup Language (KML),
- 3) Probabilistic output in METOC TIFF (MTIF),
- 4) Two-dimensional images of probabilities and input fields, and
- 5) Time series images of spatially averaged probabilities and input fields with climatological reference.

1.1.2. Product Requirements

TCFP falls under the NESDIS Level Requirement (v1.0) [REQ-001]: “NESDIS will provide environmental data, information, products, services, and reports in the Foundational,

Geophysical, and Analytical thematic product areas.”, and specifically, SPSRB request #0806-003: “NESDIS/StAR Tropical Cyclone Formation Probability product extension to the S. Pacific and Indian Oceans.” The TCFP guidance product supports the “Tropical Cyclone Characteristics” category. Table 1-1 lists the TCFP guidance product NESDIS Product Baseline requirements.

Table 1-1. NESDIS Product Baseline requirements for TCFP.

Baseline Products	Data Type	Geographic Coverage	Refresh Rate	Latency
Tropical Cyclone Formation Probability (TCFP)	Gridded	Global Ocean	6 hours	4 hours

1.2. Satellite Instrument Description

To provide a global capability, TCFP uses longwave infrared imagery centered on a water vapor absorption band available from the global constellation of geostationary satellites. Currently, digital imagery information (i.e. brightness temperatures) come from the U.S. NOAA Geostationary Operational Environmental Satellites (GOES)-R Series, Japan Meteorological Agency (JMA) third generation satellites *Himawari-8* and *Himawari-9*, and European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Meteostat Second Generation (MSG) and Meteostat Third Generation (MTG). Each series/generation of satellite from these organizations have water vapor absorption bands with differing central wavelengths and spectral response functions. TCFP calibrates these imagery sources to the Gridded Satellite (GridSat-B1) Climate Data Record water vapor central wavelength near 6.7 μm (Knapp et al. 2011).

2. ALGORITHM DESCRIPTION

2.1. Processing Outline

The TCFP guidance product follows these steps in the list below and Fig. 2-1 when generating probabilistic output of tropical cyclogenesis. The algorithm:

- 1) reads in each dataset (and calibrates if necessary),
- 2) maps the data to the TCFP grid,
- 3) performs area or azimuthal averaging of values at each grid point,
- 4) calculates predictors and derived quantities,
- 5) generates interim output files (the algorithm reuses interim output for low refresh rate data),
- 6) predicts probabilistic output using predictors, and
- 7) generates output data files and figures.

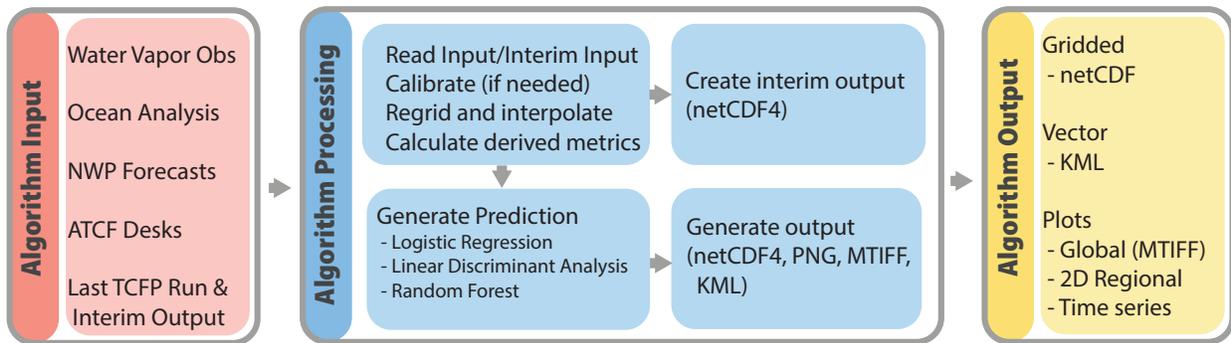


Figure 2-1. A flow diagram of TCFP processing.

2.2. Algorithm Input

The TCFP guidance product uses input data from the global constellation of geostationary satellites, global sea surface temperature products, numerical weather prediction forecasts, operational tropical cyclone characteristics from the Automated Tropical Cyclone Forecasting (ATCF) system database files (Sampson and Schrader 2000), and look up tables containing climatology and model and calibration coefficients. Table 2-1 contains a list of algorithm input with source information, spatial and temporal resolution, and description. Figure 2-2 shows the algorithm input start times relative when the algorithm run and processing create time.

TIME (UTC)	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
GFS	18 UTC Day-1				00 UTC				06 UTC				12 UTC				18 UTC							
Ocean	Day-1 SST product								Today's SST product															
Satellite	23	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22
TCFP Start	18 UTC Day-1				00 UTC				06 UTC				12 UTC				18 UTC							

Figure 2-2. A diagram depicting the algorithm input start times relative to algorithm run and processing create times. Color progresses from light to dark throughout the day.

Table 2-1. TCFP guidance product algorithm input data with source, spatial and temporal coverage, and description.

Algorithm Input	Source	Coverage		Description
		Spatial	Temporal	
Water Vapor Imagery (~6.7 μm)	NOAA GOES-R Series (GOES-16/-18)	2 km	10-min full disk	Provides global convective activity and subsidence data from longwave infrared imagery in the water vapor absorption band near 6.7 μm
	JMA Third Generation (Himawari-9)	2 km	10-min full disk	
	Meteosat Second Generation (Meteosat-9/-10/-11)	3 km	15-min full disk	
	Meteosat Third Generation (Meteosat-12)	2 km	10-min full disk	
Numerical Weather Prediction Forecasts	NOAA Global Forecast System	0.25° × 0.25°	Generated relative to synoptic hours (0000, 0600, 1200, 1800 UTC) and provides forecasts at 6 h intervals	Provides horizontal and vertical temperature, relative humidity, and wind analysis and forecast information
Sea Surface Temperature	Canadian Meteorological Center Group for High Resolution Sea Surface Temperature Level 4	0.01° × 0.01°	Daily	Provides a global blended (satellite and in situ) SST analysis
Tropical Cyclone Characteristics	NOAA National Hurricane Center, NOAA Central Pacific Hurricane Center, Dept. of Defense Joint Typhoon Warning Center from the Automated Tropical Cyclone Forecasting (ATCF) system	Variable at 0.10°	Variable but at 6 to 12 h when available	Tropical cyclone warning centers generate center position, maximum sustained wind (intensity), and development level (e.g., disturbance, tropical storm, tropical cyclone) characteristic data
Ancillary Data	TCFP Algorithm Developers	1.00° × 1.00° (if applicable)	6 h (if applicable)	Climatology for algorithm predictors and machine learning model and water vapor calibration coefficient look up tables

2.3. Theoretical Description

2.3.1. Physical Description

The tropical cyclone research community has long documented the locations and the conditions that support tropical cyclone formation (Gray 1968, 1975). These works highlight the necessary environmental conditions associated with tropical cyclone genesis (e.g., warm sea surface temperatures, low vertical wind shear, low-level cyclonic rotation). Other works have added important information related to the role of convection in determining whether a tropical disturbance will further develop into a tropical cyclone (Zehr 1992).

The TCFP guidance product captures the physical insights provided through the community into input predictors into statistical/machine learning algorithms to generate the likelihood of tropical cyclogenesis. The algorithm does this by using four data sources: 1) geostationary satellite imagery, 2) global ocean product sea surface temperatures, 3) numerical weather prediction analysis and forecast output, 4) operational tropical cyclone characteristics. The geostationary satellite longwave infrared water vapor absorption band imagery captures both the convective activity associated with potential disturbances in the tropics and the subsidence associated with the subtropical ridge. Sea surface temperatures provide insight both into the energy source needed to develop tropical deep convection and maintain a disturbance as it forms into a tropical cyclone, as well as identifying regions where formation is not possible (Palmén 1948; Dare and McBride 2011). Numerical weather prediction model analysis and forecast output provide the current and evolution of atmospheric state variables. From the state variables, the algorithm calculates large-scale environmental diagnostic metrics like vertical wind shear and vertical instability. These diagnostic quantities provide insight into the development, maintenance, and organization of tropical deep convection. Operational tropical cyclone characteristics provide locations of currently active tropical cyclones so that these systems can be removed from the product and not inappropriately flagged for genesis.

With these physical predictors the algorithm uses linear discriminate analysis, logistic regression, and random forest classifiers to produce an equally-weighted consensus for the likelihood of tropical cyclogenesis. Each consensus member has strengths and weaknesses in producing realistic probabilities of genesis. The consensus approach balances these properties. Each statistical/machine learning algorithm has two versions that contribute to the consensus: 1) a global version that uses all global information in the training process and 2) a regional version that trains only on the unique characteristics contained within a region. Tropical cyclogenesis is a rare event, which means that the sample is imbalanced to null events. The global version of each algorithm benefits from the larger sample size, but can perform poorly in regions where conditions (e.g., thermodynamics) are consistently marginal. The regional version's performance benefits from understanding these marginal conditions, but is hampered by its smaller training sample size. This global plus regional approach improves TCFP's overall performance.

2.3.2. Mathematical Description

TCFP version 4 changes the mathematical approach to calculating predictors including calculations involving area averaging, temperature gradient and advection, and vertical instability. TCFP also uses new predictors including shallow-layer vertical wind shear, generalized wind shear, and output from an entrained plume cloud model. The subsequent sections outline how the current version calculates the metrics used to create diagnostic fields and probabilistic formation output. Slocum et al. (2022) summarize the calculation of these metrics, and provide a storm centric comparison between the NOAA National Centers for Environmental Prediction Global Forecast System and the European Centre for Medium-Range Weather Forecasts fifth generation atmospheric reanalysis.

2.3.2.1. Azimuthal Averaging

Previous versions of the algorithm take an area average in either a $5^\circ \times 5^\circ$ box (Schumacher et al. 2009) or a 500 km radius circle. TCFP version 4 takes a different approach, akin to other tropical cyclone following statistical-dynamics aids such as the Statistical Hurricane Intensity Prediction Scheme (SHIPS; DeMaria and Kaplan 1994) and the associated SHIPS developmental dataset (CIRA 2024). Aids like SHIPS map data to a cylindrical grid and calculate quantities over varying areas. 0–1000-km or 0–500-km radius circles or 200–800-km annuli are typical azimuthal areas for dynamic, kinematic, and thermodynamic fields, respectively. These areas are depicted in a storm centric perspective on Hurricane Harvey (2017) in Fig. 2-3. TCFP version 4 performs these same storm-centric calculations using analysis grid point locations instead of the storm centers (i.e., Eulerian vs Lagrangian as shown Fig. 2-4).

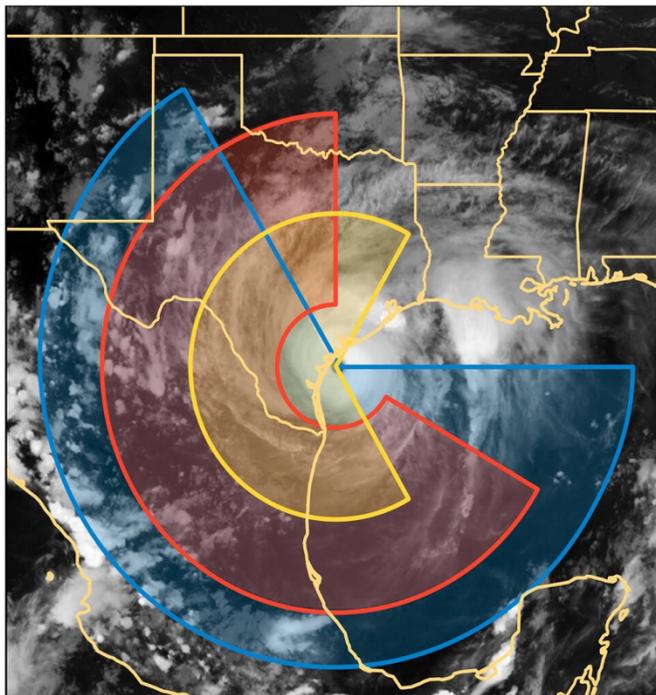


Figure 2-3. Graphic from Slocum et al. (2022) depicting *GOES-16* 10.3- μm longwave infrared observations at 0000 UTC 26 Aug 2017 showing Hurricane Harvey prior to landfall with the area-averaging regions for large-scale dynamics from 0 to 1000 km (blue circle), thermodynamics from 200 to 800 km (red annulus), and kinematics from 0 to 500 km (yellow circle). Published 2022 by the *American Meteorological Society*.

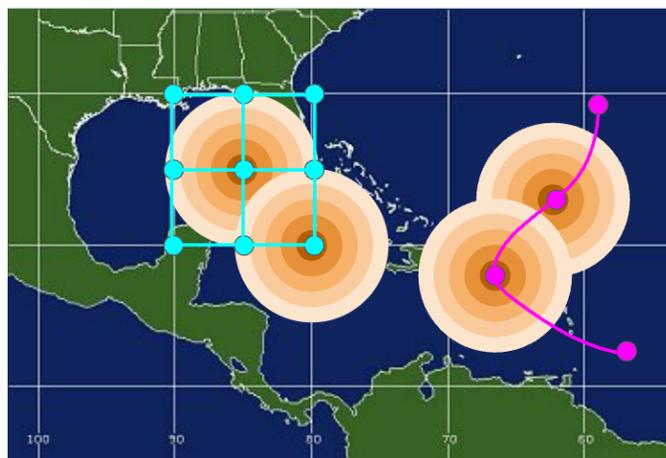


Figure 2-4. The difference between azimuthal means for TCFP as a fix grid product in cyan (i.e., Eulerian) vs a storm position-following product like SHIPS in magenta (i.e., Lagrangian). The tan rings demonstrate two cases with averaging bands 100 km in radius wide around points to show overlap.

2.3.2.2. Geostationary Satellite and Ocean Predictors**Geostationary Satellite**

The TCFP guidance product calculates simple average-based metrics for geostationary satellite predictors. Prior to calculating the metrics, TCFP intercalibrates the geostationary satellite imagery. The global constellation of geostationary satellites is inconsistent with respect to the central wavelengths and spectral response functions available in the longwave infrared water vapor absorption bands. For TCFP, the Gridded Satellite (GridSat-B1) Climate Data Record provides a baseline product to attempt to calibrate real-time imagery (Knapp et al. 2011). Using the water vapor central wavelength near 6.7 μm , one of each of the on-orbit imager's water vapor channels is regressed to 6.7 μm by solving

$$\min_w ||Xw - y||_2^2$$

and applying the correction with

$$y = Xw + C,$$

where X is the input brightness temperature at a central wavelength in a water vapor absorption band, y is the target water vapor brightness temperature near 6.7 μm , w is the slope, and C is the y intercept. With the intercalibrated brightness temperature values, TCFP calculates two input area-averaged predictors:

- 1) cloud-cleared water vapor brightness temperature within 500 km of the grid point, and
- 2) percent of pixels below -40°C within 500 km of the grid point.

For the cloud-cleared water vapor brightness temperature within 500 km of the grid point, any points less than -40°C are set to -40°C .

Ocean Predictors

As an ocean predictor, TCFP calculates the sea surface temperature averaged within 50 km of a grid point using the Canadian Meteorological Center Group for High Resolution Sea Surface Temperature Level 4. For the developmental dataset, TCFP blends the operational sea surface temperature product with the daily and weekly Optimum Interpolated Sea Surface Temperature product (Reynolds 1988).

2.3.2.3. Large-scale Environment Predictors

The product uses two forms of the large-scale environmental predictors. TCFP calculates the first form directly from the model fields using azimuthal averaging (see the top half of Table 2-2). These predictors include environmental mean sea level pressure, divergence, vorticity, vertical layer averaged relative humidity, and temperature anomaly.

TCFP derives the second form of environmental quantities using a model-based sounding. These sounding-derived predictors include layered and generalized vertical wind shear, thermal wind, vertical instability, and entrained plume cloud model output (e.g., convective inhibition, mass-weighted vertical velocity). To provide representative

examples, TCFP sample derivations use Table 2-3, which contains the mean Tropical North Atlantic Ocean atmospheric sounding from Dunion (2011).

Table 2-2. TCFP large-scale environment predictors.

Predictor	Units	Type	Description
Mean Sea Level Pressure	hPa	Calculated	200–800-km average mean sea level pressure
Divergence	s ⁻¹	Calculated	0–1000-km average 200-hPa divergence
Vorticity	s ⁻¹	Calculated	0–1000-km average 200- and 850-hPa vorticity
Temperature Anomaly	°C	Derived	0–100-km, 400–300-hPa temperature anomaly with respect to 1500-km mean temperature
Layered Vertical Wind Shear	kt	Derived	The wind difference between two levels that is defined as the 0–500-km radius 850–200-hPa deep-layer shear and 850–500-hPa shallow-layer shear
Generalized Vertical Wind Shear	kt	Derived	The 850–200-hPa mass-weighted deviation from the mean wind profile calculated from 0–500-km radius
Temperature Gradient	°C m ⁻¹	Derived	The magnitude of the temperature gradient between 850 and 700 hPa averaged from 0 to 500 km estimated from the geostrophic thermal wind
Temperature Advection	°C s ⁻¹	Derived	The temperature advection between 850 and 700 hPa averaged from 0 to 500 km estimated from the geostrophic thermal wind
Vertical Instability	°C	Derived	The mass-weighted summation from the surface to the level of neutral buoyancy of the difference between the surface equivalent potential temperature and saturated equivalent potential temperature
Convective Inhibition	J kg ⁻¹	Derived	The energy preventing an air parcel from rising from the surface to the level of free convection calculated from the entrained plume cloud model
Average Vertical Velocity	m s ⁻¹	Derived	The average vertical velocity of an air parcel from rising from the surface to the level of neutral buoyancy calculated from the entrained plume cloud model

Table 2-3. Mean Tropical North Atlantic Ocean atmospheric sounding of pressure, temperature, relative humidity, wind speed, and wind direction adapted from Table 1 in Dunion (2011). Published 2011 by the *American Meteorological Society*.

Pressure (hPa)	Temperature (°C)	Relative Humidity (%)	Wind Speed (kt)	Direction (°)
50	-63.1	25.4	25.0	88
100	-74.4	32.9	7.8	70
150	-67.0	33.9	6.6	321
200	-54.4	34.2	7.4	304
250	-42.6	34.8	4.7	302
300	-32.6	34.5	2.1	313
400	-17.3	37.5	2.3	81
500	-6.5	41.7	4.5	93
600	1.7	48.8	6.2	100
700	9.1	54.4	8.3	100
850	17.4	76.4	10.3	101
925	21.7	83.2	10.5	100
1000	26.4	81.4	6.4	92
1015.3 (Surface)	26.9	81.3	3.9	91

Temperature anomaly

Tropical and subtropical cyclones exhibit symmetric or asymmetric warm-core anomalies through the middle to upper troposphere that result from the vortex structure and convection, where the temperature difference between some central point and the surrounding environment represent the warm core. As cyclones strengthen, the warm-core anomaly increases in magnitude relative to the surrounding environment and expands in size. Frank (1977) shows that the temperature perturbation from the tropical cyclone extends 1500 km from the center of the tropical cyclone. To calculate the warm-core anomaly, TCFP subtracts the 1500-km mean 400–300 hPa temperature profile from the 0–100-km, 400–300-hPa temperature profile to calculate a 400–300-hPa temperature anomaly.

Vertical Wind Shear

The change in wind speed and direction as a function of height can hinder the organization of deep convection. To measure the changes in the profile, vertical wind shear metrics capture these in the vertical wind profile. TCFP calculates two forms of

vertical wind shear. The first is the classic layered shear in which the magnitude of one layer is subtracted from another. The algorithm calculates two layered wind shears:

- 1) deep-layer shear from 200–850 hPa and
- 2) shallow-layer shear from 500–850 hPa.

The formula for layered shear is

$$SHR = [(u_t - u_b)^2 + (v_t - v_b)^2]^{1/2},$$

where u is the zonal wind component, v is the meridional wind component, subscript t is the top of the layer (i.e., 200 or 500 hPa), and subscript b is the bottom of the layer (i.e., 850 hPa).

The second is the generalized wind shear (Knaff et al. 2005), which TCFP calculates using

$$G = 4 \sum_{i=1}^I w_p [(u_p - \bar{u})^2 + (v_p - \bar{v})^2]^{1/2},$$

where i is the pressure level index for pressure values p from 850 to 200 hPa, I is the total number of pressure levels, w is the mass weight for the pressure level p , and \bar{u} and \bar{v} are the column-averaged winds.

Figure 2-5 shows a hodograph of the vertical wind profile from the mean Tropical North Atlantic Ocean sounding in Table 2-3 (blue curve) and the vectors for the deep-layer shear (red), shallow-layer shear (orange), and mean wind (yellow). From Table 2-3, TCFP would calculate 17.3, 5.9, and 17.6 kt as the values of deep-layer, shallow-layer, and generalized vertical wind shear, respectively.

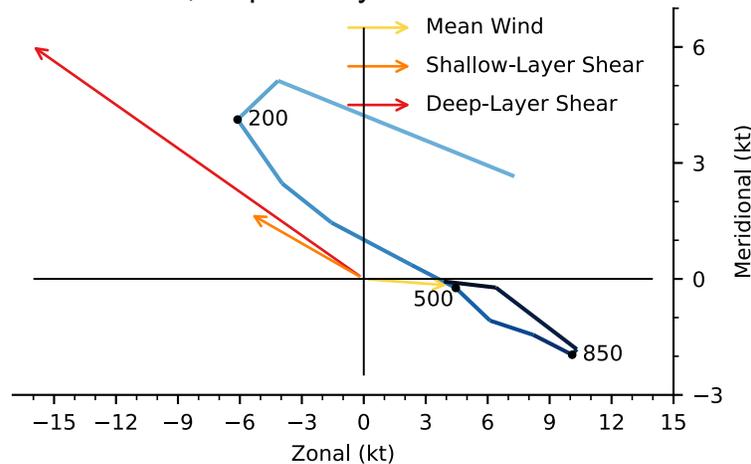


Figure 2-5. A hodograph depicting the vertical wind profile from 1015 to 100 hPa (blue-shaded curve and black dots) and vectors for the 850–200-hPa deep-layer shear (red), 850–500-hPa shallow-layer shear (orange), and 850–200-hPa mass-weighted mean wind (yellow).

Temperature Gradient and Advection

TCFP version 4 updated the approach to calculating temperature gradient and advection. TCFP uses the geostrophic thermal wind equation to calculate the temperature gradient and advection where each takes the form

$$\begin{aligned} \text{TADV} &= -\bar{u} \frac{\partial T}{\partial x} - \bar{v} \frac{\partial T}{\partial y} \\ \text{TGRD} &= \left[\left(\frac{\partial T}{\partial x} \right)^2 + \left(\frac{\partial T}{\partial y} \right)^2 \right]^{1/2}, \end{aligned}$$

where \bar{u} and \bar{v} are the average 700- and 850-hPa winds and

$$\begin{aligned} \frac{\partial T}{\partial x} &= -\frac{f}{R_0} \frac{\partial v}{\partial \ln p} \\ \frac{\partial T}{\partial y} &= -\frac{f}{R_0} \frac{\partial u}{\partial \ln p}, \end{aligned}$$

where f is the Coriolis parameter as a function of TCFP grid point and R_0 is the universal gas constant.

Vertical Instability

TCFP uses several metrics to understand the potential to develop and maintain deep convection. The vertical instability metric is one example. To calculate vertical instability, TCFP computes the surface equivalent potential temperature and the saturated equivalent potential temperature with the formulas in Bolton (1980). Then, TCFP calculates a summation of the difference between the surface equivalent potential temperature and the saturated equivalent potential temperature to yield the vertical instability as defined by

$$S = \sum_{i=1}^I w_p (\theta_{e, \text{SFC}} - \theta_{e,p}^*),$$

where i is the pressure level index for pressure values p from the surface (SFC) to the level of neutral buoyancy (LNB), w is the mass weight for the pressure level p , $\theta_{e, \text{SFC}}$ is the surface equivalent potential temperature, $\theta_{e,p}^*$ is the saturated equivalent potential temperature profile. Note that S is negative from the surface to the level of free convection and positive from the level of free convection to the level of neutral buoyancy.

Using the vertical profile of temperature along with the surface relative humidity listed in Table 2-3 as an example, TCFP calculates the vertical instability as follows. TCFP determines the gray area shown in Fig. 2-6 located between the surface to the level of neutral buoyancy by using the difference between the surface equivalent potential temperature ($\theta_{e, \text{SFC}}$), the black line, and vertical profile of saturated equivalent potential temperature (θ_e^*), the blue line. In this example as shown in Fig. 2-6, TCFP would calculate a vertical instability value of 6.3 °C.

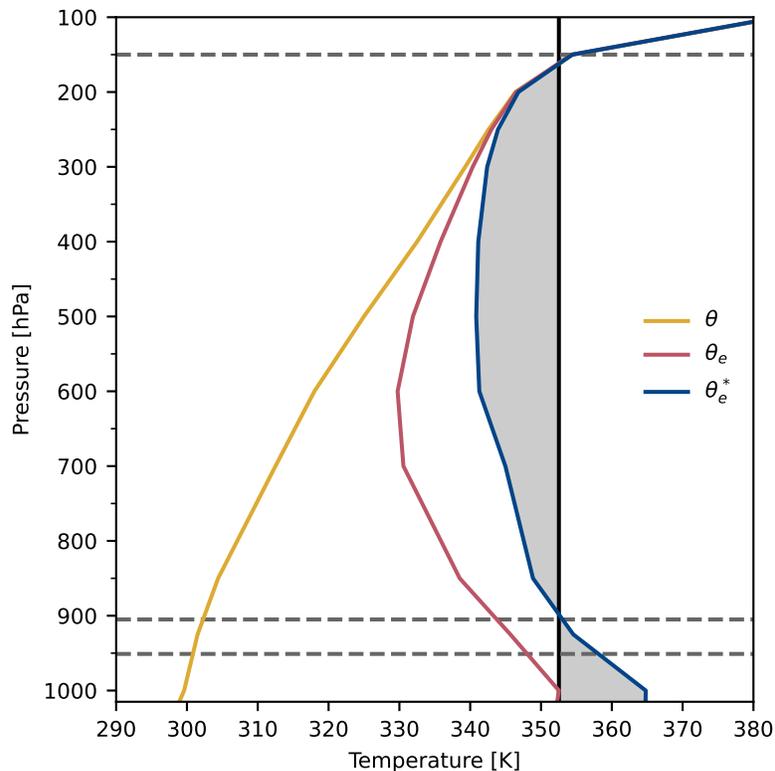


Figure 2-6. Vertical profiles of potential temperature (yellow curve), equivalent potential temperature (red curve), and saturated equivalent potential temperature (blue curve) based on the mean Tropical North Atlantic Ocean atmospheric sounding in Table 2-2. The black line indicates the value of the surface equivalent potential temperature ($\theta_{e, SFC}$). The gray shaded area is the area used to calculate the vertical instability predictor. The gray dashed lines from bottom to top represent the lifted condensation level, level of free convection, and level of neutral buoyancy.

Entrained Plume Cloud Model

Vertical instability is one measure necessary for understanding the conduciveness of the environment for developing and maintaining deep convection. TCFP employs other measures as well, namely, metrics from a parcel-based, Lagrangian cloud model (DeMaria 2009) used in tropical cyclone following statistical–dynamical aids. DeMaria (2009) bases the cloud model on the Simpson and Wiggert (1969) entraining plume model, but adds the thermodynamics and bulk microphysics of Ooyama (1990), which allow accounting for the water and ice phase condensate parcel weighting. TCFP uses model-based profiles of temperature and relative humidity as input into the cloud model. TCFP assumes the following assumptions on air parcel’s initial conditions: 15 m s⁻¹ initial parcel vertical velocity [consistent with SHIPS (CIRA 2024), but an increase from the 8 m s⁻¹ used in DeMaria (2009)], 500 m parcel radius, 0.1 nondimensional entrainment

coefficient, and $1/600 \text{ s}^{-1}$ precipitation removal rate. From these assumptions, Fig. 2-7 shows the environmental temperature from Table 2-3 along with the cloud model calculated parcel virtual potential temperature and parcel vertical velocity as a function of pressure. For the example sounding (Table 2-3), TCFP calculates an average vertical velocity value of 6.6 m s^{-1} and a convective inhibition value of 29.8 J kg^{-1} .

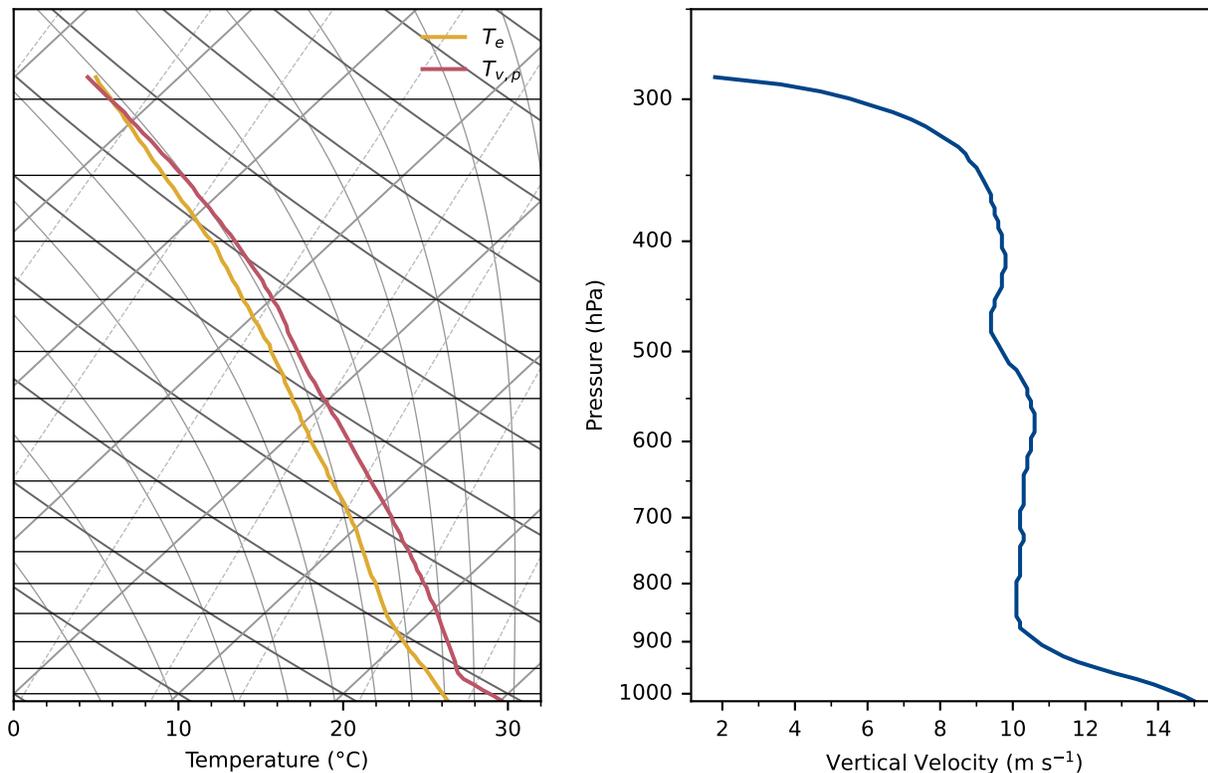


Figure 2-7. Vertical profiles of the (left) environmental temperature (yellow curve) and parcel virtual temperature (red curve) on a skew- T log- p diagram and (right) parcel vertical velocity as calculated by the entrained plume cloud model.

2.3.2.4. Predicting Probabilities

TCFP uses an equally-weighted consensus of probabilistic statistical/machine learning algorithms that are trained with global and regional versions (Table 2-4). Here, TCFP employs three algorithms to estimate probabilities, specifically, linear discriminant analysis, logistic regression, and random forest classifiers. For training, the TCFP algorithm uses the implementations of the statistical/machine learning algorithms available through scikit-learn (Pedregosa et al. 2011) and generates output using Fortran-based drivers for the algorithms.

Linear Discriminant Analysis

Linear Discriminant Analysis is a dimensionality reduction technique applicable to classification problems. LDA strives to use predictor information to maximize class separation while projecting the predictor onto a lower dimensional. With the calculated weights, the probability of a class is determined with

$$\log p(y_i = k|X_i) = w_k^t X_i + w_{k0} + C,$$

where $y_i = k$ is the sample classification ($k = 1$ for a 'yes' tropical cyclone genesis event), X is a vector of input features, w are the vector of weights for each input feature, and C is a constant.

Logistic Regression

Logistic regression attempts to generate the log-odds of an event as a linear combination of predictor information. Because of its formulation, logistic regression generates the true likelihood of an event unlike other machine learning or statistical methods that generate the proportion of a training sample that is similar to an event's input predictor information. To generate the probability of a class, the following is used:

$$p(y_i = k|X_i) = \frac{1}{1 + \exp(-X_i w_k - w_{k0})},$$

where $y_i = k$ is the sample classification ($k = 1$ for a 'yes' tropical cyclone genesis event), X is a vector of input features, w are the vector of weights for each input feature.

Random Forest

The random forest algorithm (Breiman 2001) is an ensemble of decision trees where each decision tree may see a random subset of input predictor or subset of the training sample to increase the variance of the ensemble. Decision trees attempt to split the training samples into classifications ('yes' tropical cyclone genesis or 'no' genesis) using the information. In making the split, the tree uses Gini impurity where impurity is measured as how "clean" the split is in correctly assigning 'yes' and 'no' cases to a new group. The split occurs at a node. The tree reaches terminal node either by an algorithm setting or when only a single sample remains in the last node.

The probabilistic from the random forest is the mean of the proportion of samples representing the class k at the terminal mode in the individual decision trees.

$$p(y_i = k|X_i) = \frac{1}{M} \sum_{m=1}^M \frac{1}{n_m} \sum_{y \in X_m} I(y = k),$$

where $y_i = k$ is the sample classification ($k = 1$ for a 'yes' tropical cyclone genesis event), X is a vector of input features, M is the total number of decision trees, m is the decision tree terminal node, n is the number of samples in the node, and I is the input training data residing at the terminal node.

Calibration

Before constructing the equally weighted consensus, the probabilistic output needs to be calibrated so that the output from each statistical/machine learning method represents the true likelihood of the event rather than the proportion of the training sample (e.g., random forest). After calibration, the individual consensus members' predicted probability values align more closely with the observed relative frequency of the event [this is often visualized with calibration curves, reliability diagrams, or attribute diagrams (for details, see Wilks 2019)].

Here, TCFP uses sigmoid regression [also called Platt's logistic model (Platt 2000)] to calibrate model probabilities. The below equation defines the Platt's logistic model:

$$p(y_i = 1|f_i) = \frac{1}{1 + \exp(Af_i + B)}$$

where f is the algorithm output probability, A and B are the sigmoid regression coefficients, and $p(y = 1|f)$ is the calibrated probability.

Subregion Formation Probability

For the time series output, TCFP calculates area weighted averages of quantities within subregions (discussed in section 2.4). For probabilities, this should not be done as a straight average when determining the probability of tropical cyclone formation within the subregion. Instead, TCFP uses the Additive Law of Probability where the union of the set follows $p(A \text{ or } B) = p(A) + p(B) - p(A \text{ and } B)$, where $p(A \text{ and } B) = p(A) \times p(B)$ is nonzero. To expand this beyond the probability of two points, A and B , to all points in the subregion, TCFP calculates the probability as

$$p_s = \prod_{i=1}^I [1 - w_i p(i)],$$

where p_s is the probability of formation in a subregion, $p(i)$ is the probability of formation at a TCFP domain grid point within the subregion, w_i is the area weight based on the cosine of latitude, and I is the total number of points within the subregion.

2.4. Algorithm Output

TCFP generates five types of output: 1) probabilistic output in Network Common Data Form (NetCDF), 2) probabilistic output in Keyhole Markup Language (KML), 3) probabilistic output in METOC TIFF (MTIF), 4) two-dimensional images of probabilities and input fields, 5) time series images of spatially averaged probabilities and input fields with climatological reference. For probabilistic output, note that data points with active tropical cyclones are removed using tropical cyclone characteristic information from the NOAA National Hurricane Center, NOAA Central Pacific Hurricane Center, and Department of Defense Joint Typhoon Warning Center.

On the landing page for the real-time TCFP product, TCFP displays global formation probability as shown in Fig. 2-8. For the two-dimensional images of probabilities, input

fields, and time series of spatially averaged probabilities and input fields with climatological reference, TCFP divides the global oceans into large-area regions. Examples are provided in Fig. 2-9, and subregions, Fig. 2-10, within each large area (see Table 2-4). Both options allow product users to diagnose conditions impacting tropical cyclone formation and better cater to the user's individual needs. For the physical fields in the two-dimensional and time series plots, TCFP displays nine quantities as follows:

- 1) cloud-cleared water vapor brightness temperature (Fig. 2-11),
- 2) water vapor percent pixel below than -40°C (Fig. 2-12),
- 3) sea surface temperature (Fig. 2-13),
- 4) mean sea level pressure (Fig. 2-14),
- 5) 850-hPa lower-tropospheric vorticity (Fig. 2-15),
- 6) 850-hPa lower-tropospheric divergence (Fig. 2-16),
- 7) generalized vertical wind shear (Fig. 2-17),
- 8) vertical instability (Fig. 2-18), and
- 9) thermal wind temperature advection (Fig. 2-19).

TCFP provides three types of two-dimensional plots for each physical field. These include the current values (averaged over the appropriate period, e.g., 0 to 24 h and 24 to 48 h), the two-week rolling mean of the climatological values, and the anomalies between the current and climatological values. On the time series plots, TCFP shows the 6-hourly values as a scatter plot in time, a 14-day running mean of the values as a curve (this dampens some of the diurnal variability in predictors like the vertical instability), and climatological values at the 10th, 30th, 50th, 70th, 90th percentiles of the two-week running mean as gray shading.

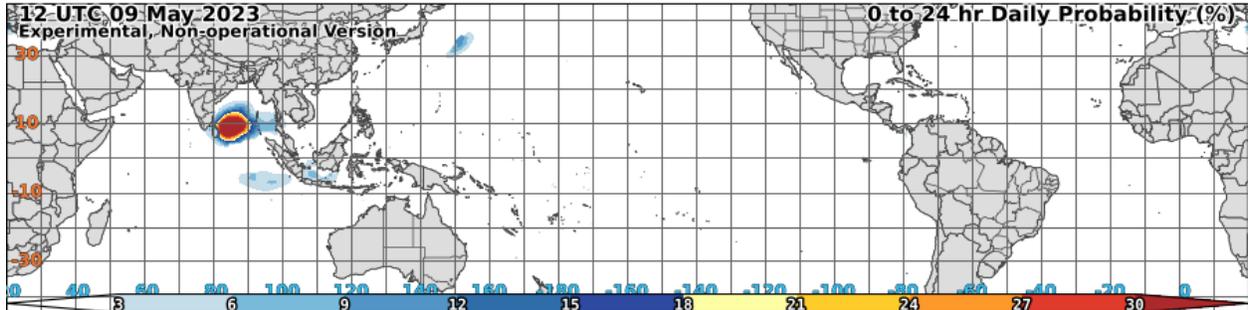


Figure 2-8. Global 0 to 24 h TCFP guidance product output tropical cyclogenesis probabilities on 1200 UTC 9 May 2023.

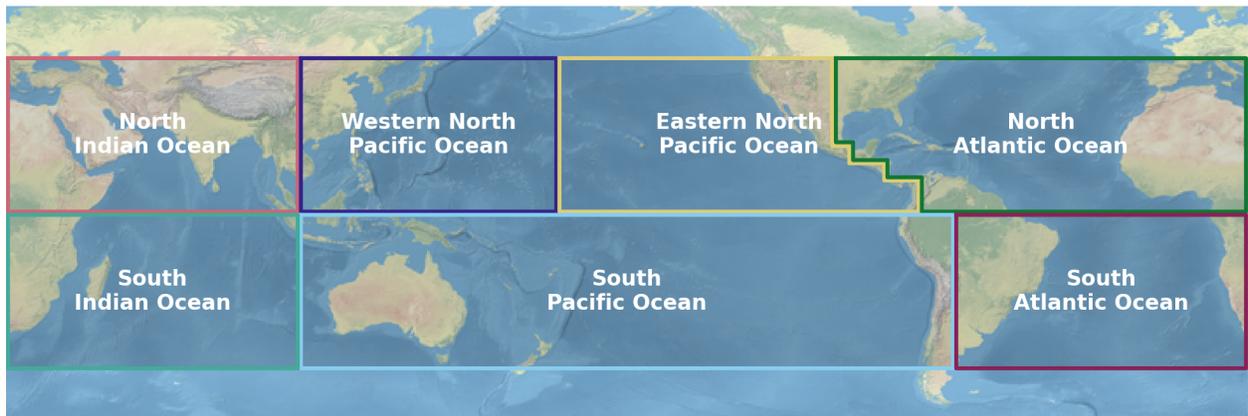


Figure 2-9. TCFP domain averaging large-scale regions.

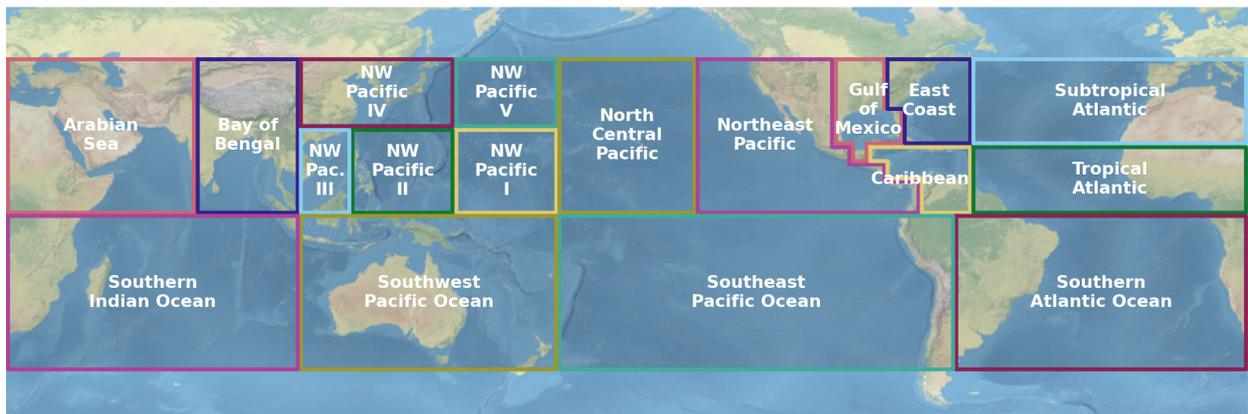


Figure 2-10. TCFP domain averaging subregions.

Table 2-4. TCFP coefficient and averaging region and subregion names along with general longitudinal and latitudinal extent.

Name	Type	Region	General Geographical Extent			
			Longitudinal		Latitudinal	
			West	East	North	South
North Atlantic	Region	—	100°W	20°E	45°N	0°
South Atlantic	Region	—	65°W	20°E	0°	45°S
Eastern North Pacific	Region	—	180°	75°W	45°N	0°
Western North Pacific	Region	—	105°E	180°	45°N	0°
North Indian	Region	—	20°E	105°E	45°N	0°
South Indian	Region	—	20°E	105°E	0°	45°S
Southern Pacific	Region	—	105°E	65°W	0°	45°S
Tropical Atlantic	Subregion	North Atlantic	60°W	20°E	45°N	20°N
Caribbean	Subregion	North Atlantic	90°W	60°W	20°N	0°
Gulf of Mexico	Subregion	North Atlantic	100°W	80°W	45°N	15°N
East Coast	Subregion	North Atlantic	85°W	60°W	45°N	20°N
Subtropical Atlantic	Subregion	North Atlantic	60°W	20°E	20°N	0°
Southern Atlantic	Subregion	South Atlantic	65°W	20°E	0°	45°S
Northeast Pacific	Subregion	Eastern North Pacific	140°W	75°W	45°N	0°
North Central Pacific	Subregion	Eastern North Pacific	180°	140°W	45°N	0°
Western North Pacific I	Subregion	Western North Pacific	150°E	180°	25°N	0°
Western North Pacific II	Subregion	Western North Pacific	120°E	150°E	25°N	0°
Western North Pacific III	Subregion	Western North Pacific	105°E	120°E	25°N	0°
Western North Pacific IV	Subregion	Western North Pacific	105°E	150°E	45°N	25°N
Western North Pacific V	Subregion	Western North Pacific	150°E	180°	45°N	25°N
Southeast Pacific	Subregion	Southern Pacific	180°	65°W	0°	45°S
Southwest Pacific	Subregion	Southern Pacific	105°E	180°	0°	45°S
Bay of Bengal	Subregion	North Indian	75°E	105°E	45°N	0°
Arabian Sea	Subregion	North Indian	20°E	75°E	45°N	0°
Southern Indian	Subregion	South Indian	20°E	105°E	0°	45°S

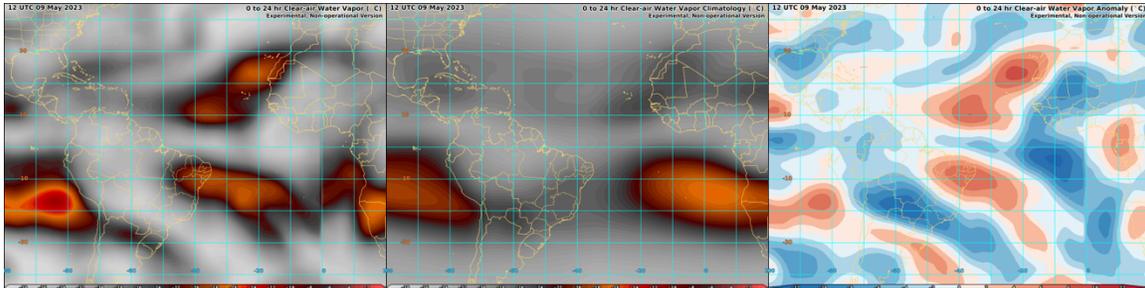


Figure 2-11. 0- to 500-km averaged cloud-cleared water vapor brightness temperature centered over the Atlantic Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.

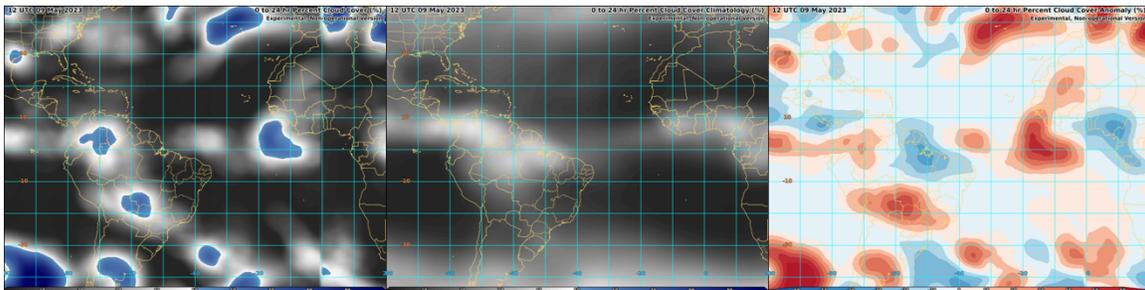


Figure 2-12. 0- to 500-km averaged water vapor percent pixel below -40°C centered over the Atlantic Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.

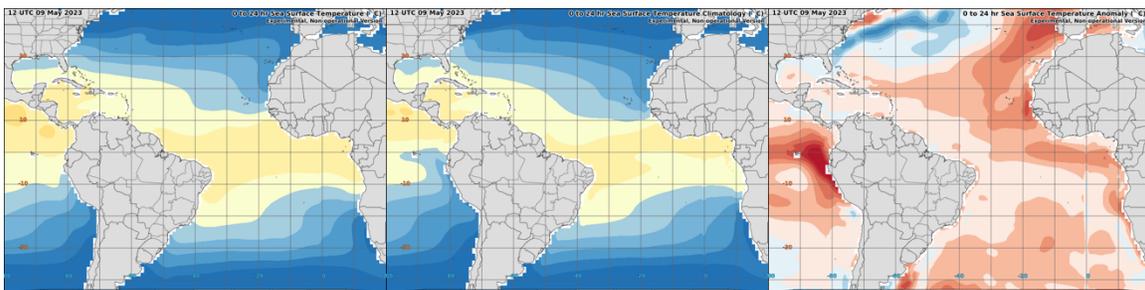


Figure 2-13. 0- to 50-km averaged sea surface temperature centered over the Atlantic Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.

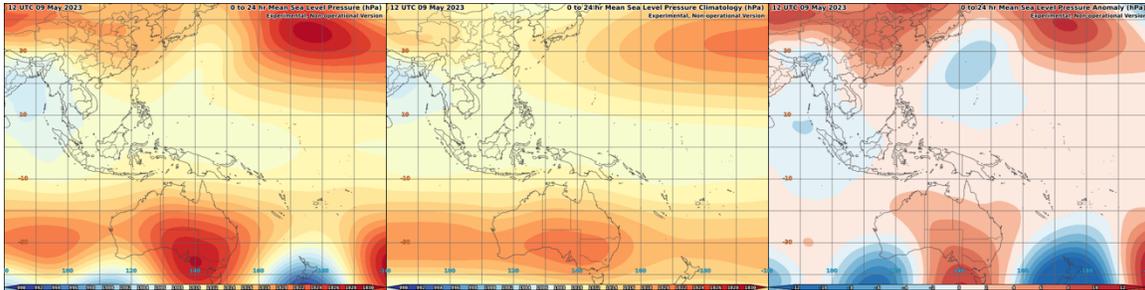


Figure 2-14. 200- to 800-km averaged environmental mean sea level pressure 0- to 50-km averaged sea surface temperature centered over the Western Pacific Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.

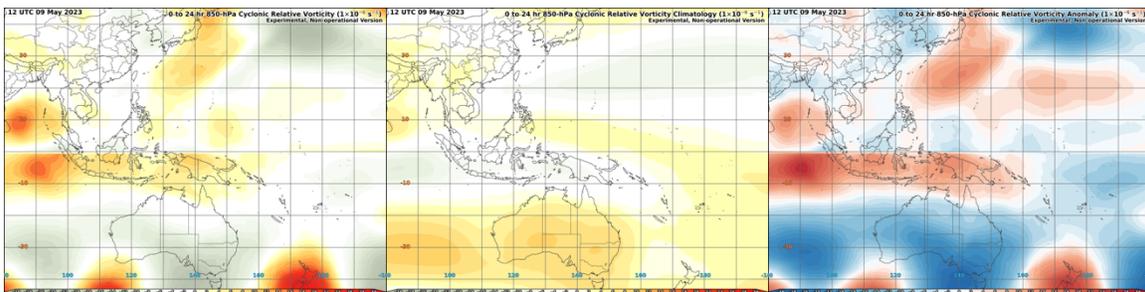


Figure 2-15. 0- to 1000-km averaged 850-hPa lower-tropospheric vorticity centered over the Western Pacific Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.

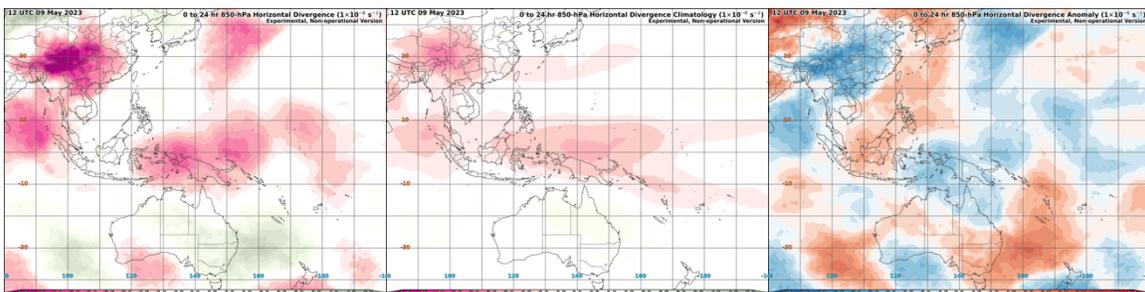


Figure 2-16. 0- to 1000-km averaged 850-hPa lower-tropospheric divergence centered over the Western Pacific Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.

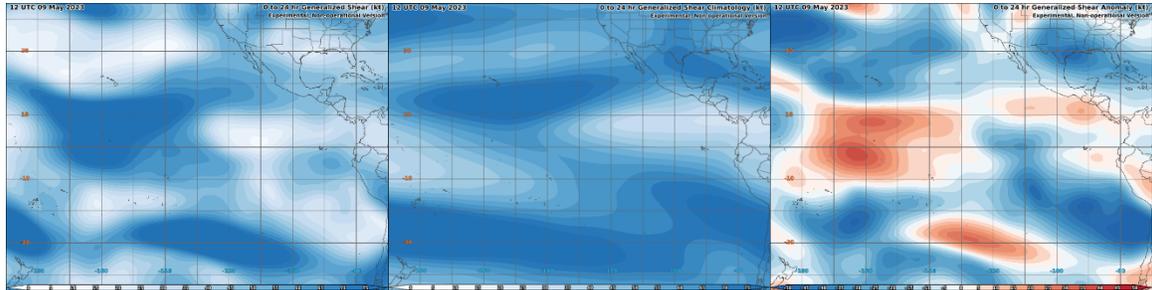


Figure 2-17. 0- to 500-km generalized vertical wind shear centered over the Eastern and Central Pacific Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.

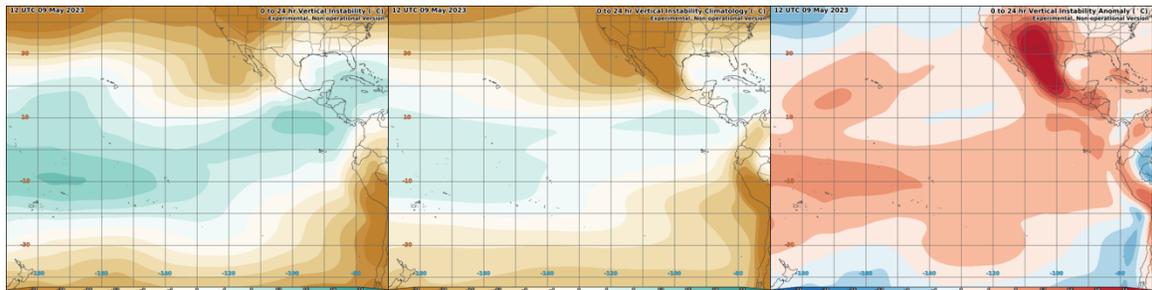


Figure 2-18. 200- to 800-km vertical instability centered over the Eastern and Central Pacific Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.

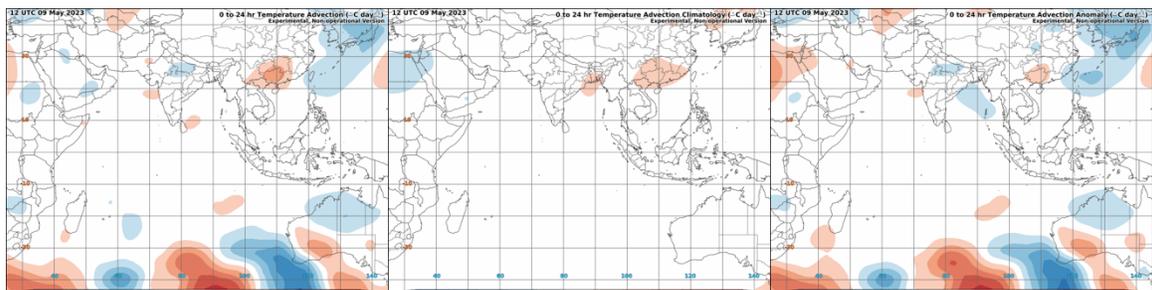


Figure 2-19. 0- to 500-km 850-hPa thermal wind temperature advection centered over the Indian Ocean for (left) real time, (center) climatology, and (right) anomaly on 1200 UTC 9 May 2023.

2.5. Performance Estimates

2.5.1. Test Data Description

TCFP includes unit and integration tests for the algorithm. The unit tests cover calculating each physical parameter and use validation data from the mean Tropical North Atlantic sounding (Table 2-3) and validation from the literature (e.g., Bolton 1980). These unit tests cover a variety of conditions and extremes to ensure expected algorithm output. Integration tests include sample cases from the TCFP developmental dataset to make sure that the algorithm creates consistent output between multiple computational environments.

2.5.2. Sensor Effects

The current version does not take into account sensor effects such as limb corrections or solar corrections. Outside of overlap, any sensor effects (e.g., limb effects, striping, stray light) may negatively impact the probabilistic output.

2.5.3. Retrieval Errors

The current version does not handle retrieval errors contained in the radiance data from the water vapor absorption band longwave infrared observations from the global constellation of geostationary satellites. Any of these errors will propagate to the output from the TCFP product. Internally, errors are related to the training of the individual statistical/machine learning models within the algorithm. These errors are expected to remain consistent with those in the training process.

2.6. Practical Considerations

2.6.1. Numerical Computation Considerations

The TCFP algorithm uses the Python programming language as the algorithm driver and for algorithm input and output. From the Python driver, TCFP calls shared object libraries for the azimuthal area averages, calculating derived predictors, and statistical/machine learning model predictions, which are coded in the Fortran 2018 standard (ISO 2018). Both the Python driver and Fortran-generated shared object libraries use aspects of parallel programming. Python uses task parallelism in processing various input files simultaneously. The Fortran routines use Open Multi-Processing (OpenMP) auto parallelization features in the Fortran 2018 standard as implemented by the GNU Compiler Collection. At each step of processing the TCFP algorithm generates interim output. TCFP uses these interim output files to avoid duplicative calculations on the same fields (e.g., sea surface temperature and model) and can be used for diagnostic purposes.

2.6.2. Programming and Procedural Considerations

The Delivered Algorithm Package contains details about implementation with respect to compiler, programming language, and third-party community package versions. The algorithm is built on a new system using CMake to build the Fortran 2018 routines into a shared object library that can then be called by the Python driver scripts. Data input and output is controlled using a combination of environmental variables and configuration files in the Tom's Obvious, Minimal Language (TOML) configuration file format.

2.6.3. Quality Assessment and Diagnostics

Since TCFP is a derived product, the input data quality assurance ensures the product output quality. The algorithm has unit and integration tests to ensure that if the input data are of expected quality that the output should fall within expected parameters and behaviors. For internal purposes, the algorithm has multiple verbosity modes to monitor all aspects of the algorithm processing. Externally, the algorithm contains quality flags to highlight potential issues related to input data.

2.6.4. Exception Handling

TCFP raises all exceptions using the Python programming languages built-in exceptions. In the case that an error occurs in the Fortran-based shared object library the Fortran will return bad data values for the Python driver to interpret and raise the appropriate errors.

2.7. Validation

TCFP defines tropical cyclogenesis as a 0 to 500 km radius region around the formation location in the ATCF tropical cyclone characteristics best-track database files (Sampson and Schrader 2000) provided by the NOAA National Hurricane Center, NOAA Central Pacific Hurricane Center, and Department of Defense Joint Typhoon Hurricane Center. This 0 to 500 km radius region counts as a 'yes' and any TCFP guidance product domain point that falls within this circle is considered as a 'yes' for TC formation (>34 kt). The TCFP guidance product output grid point probabilities can then be evaluated using various verification statistics.

The first is the Brier score. The Brier score is to classification forecasts as the mean square error is to regression forecasts and is defined as

$$BS = \frac{1}{n} \sum_{i=0}^n (y_i - p_i)^2,$$

where n is the number of forecasts (both in time and on the TCFP grid), y_i is the verified 'yes' forecast on the TCFP grid, and p_i is the probabilistic output from the TCFP algorithm. Brier score is a loss metric where smaller is better.

The second approach is to convert the probabilistic forecast to a nonprobabilistic forecast by setting an optimal probability threshold as a binary ‘yes’ forecast (Wilks 2019). In nonprobabilistic form, TCFP can be evaluated using performance metrics from a 2×2 contingency table or confusion matrix where a represents the true, b the false positives, c the false negatives, and d the true. For rare events, the Peirce skill score is a useful metric, which is defined as

$$PSS = H - F,$$

where $H = a/(a + c)$ is the hit rate and $F = b/(b + d)$ is the false alarm rate (not to be confused with the false alarm ratio). See Wilks (2019) for a discussion on the Peirce skill score and other contingency table metrics.

3. ASSUMPTIONS AND LIMITATIONS

3.1. Performance Assumptions

TCFP assumes that algorithm input data is of consistent quality and timeliness. Degradations to input data either due to satellite instruments needing recalibration or experiencing hardware issues (e.g., *GOES-17* loop heat pipe subsystem issue) and increased latency causing a reliance on old data will cascade to TCFP. Errors in model output such as poorly analyzed or forecast large-scale wind fields will directly impact vertical wind shear calculations and impact output probabilities. TCFP does include quality flags to warn of some potential issues.

3.2. Potential Improvements

Tropical cyclone warning centers desire skillful formation probability information out to seven days. Future versions of TCFP will attempt to generate skillful guidance beyond the current 48 h product to provide guidance at longer forecast lead times.

4. REFERENCES

- Bolton, D., 1980: The Computation of Equivalent Potential Temperature. *Mon. Wea. Rev.*, **108**, 1046–1053, [https://doi.org/10.1175/1520-0493\(1980\)108<1046:TCOEPT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<1046:TCOEPT>2.0.CO;2).
- Breiman, L., 2001: Random forests. *Machine Learning*, **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- CIRA, 2024: SHIPS Developmental Dataset. Cooperative Institute for Research in the Atmosphere, accessed 8 Jan 2024, https://rammb.cira.colostate.edu/research/tropical_cyclones/ships/.
- Dare, R. A., and J. L. McBride, 2011: The Threshold Sea Surface Temperature Condition for Tropical Cyclogenesis. *J. Climate*, **24**, 4570–4576, <https://doi.org/10.1175/JCLI-D-10-05006.1>.
- DeMaria, M., 2009: A simplified dynamical system for tropical cyclone intensity prediction. *Mon. Wea. Rev.*, **137**, 68–82, <https://doi.org/10.1175/2008MWR2513.1>.
- DeMaria, M., and J. Kaplan, 1994: A Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic Basin. *Wea. Forecasting*, **9**, 209–220, [https://doi.org/10.1175/1520-0434\(1994\)009<0209:ASHIPS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0209:ASHIPS>2.0.CO;2).
- DeMaria, M., J. A. Knaff, and B. H. Connell, 2001: A Tropical Cyclone Genesis Parameter for the Tropical Atlantic. *Wea. Forecasting*, **16**, 219–233, [https://doi.org/10.1175/1520-0434\(2001\)016<0219:ATCGPF>2.0.CO;2](https://doi.org/10.1175/1520-0434(2001)016<0219:ATCGPF>2.0.CO;2).
- Dunion, J. P., 2011: Rewriting the Climatology of the Tropical North Atlantic and Caribbean Sea Atmosphere. *J. Climate*, **24**, 893–908, <https://doi.org/10.1175/2010JCLI3496.1>.
- Frank, W. M., 1977: The structure and energetics of the tropical cyclone I. Storm structure. *Mon. Wea. Rev.*, **105**, 1119–1135, [https://doi.org/10.1175/1520-0493\(1977\)105<1119:TSAEOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1977)105<1119:TSAEOT>2.0.CO;2).
- Gentges, D., 2019: MTIF Specification. Naval Research Laboratory, accessed 8 Jan 2024, https://www.nrlmry.navy.mil/atcf_web/docs/pdf/MTIFSpec.pdf.
- Gray, W. M., 1968: Global View of the Origin of Tropical Disturbances and Storms. *Mon. Wea. Rev.*, **96**, 669–700, [https://doi.org/10.1175/1520-0493\(1968\)096<0669:GVOTOO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1968)096<0669:GVOTOO>2.0.CO;2).
- Gray, W. M., 1975: Tropical cyclone genesis. Colorado State University Atmospheric Science Paper 234, 127 pp. https://mountainscholar.org/bitstream/handle/10217/247/0234_Bluebook.pdf.
- ISO, 2018: Fortran 2018, ISO/IEC 1539-1:2018, 630 pp. <https://www.iso.org/standard/72320.html>
- Knaff, J. A., C. R. Sampson, and M. DeMaria, 2005: An operational statistical typhoon intensity prediction scheme for the western North Pacific. *Wea. Forecasting*, **20**, 688–699, <https://doi.org/10.1175/WAF863.1>.

-
- Knapp, K. R., and Coauthors, 2011: Globally Gridded Satellite Observations for Climate Studies. *Bull. Amer. Meteor. Soc.*, **92**, 893–907, <https://doi.org/10.1175/2011BAMS3039.1>.
- Ooyama, K. V., 1990: A thermodynamic foundation for modeling the moist atmosphere. *J. Atmos. Sci.*, **47**, 2580–2593, [https://doi.org/10.1175/1520-0469\(1990\)047<2580:ATFFMT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1990)047<2580:ATFFMT>2.0.CO;2).
- Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.
- Palmén, E., 1948: On the formation and structure of tropical hurricanes. *Geophysica*, **3**, 26–38.
- Platt, J., 2000: Probabilities for SV machines. *Advances in Large-Margin Classifiers*, A. J. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, Eds., MIT Press, 61–73, <https://doi.org/10.7551/mitpress/1113.003.0008>.
- Reynolds, R. W., 1988: A Real-Time Global Sea Surface Temperature Analysis. *J. Climate*, **1**, 75–87, [https://doi.org/10.1175/1520-0442\(1988\)001<0075:ARTGSS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1988)001<0075:ARTGSS>2.0.CO;2).
- Sampson, C. R., and A. J. Schrader, 2000: The Automated Tropical Cyclone Forecasting System (Version 3.2). *Bull. Amer. Meteor. Soc.*, **81**, 1231–1240, [https://doi.org/10.1175/1520-0477\(2000\)081<1231:TATCFS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<1231:TATCFS>2.3.CO;2).
- Schumacher, A. B., M. DeMaria, and J. A. Knaff, 2009: Objective Estimation of the 24-h Probability of Tropical Cyclone Formation. *Wea. Forecasting*, **24**, 456–471, <https://doi.org/10.1175/2008WAF2007109.1>.
- Simpson, J., and V. Wiggert, 1969: Models of precipitating cumulus towers. *Mon. Wea. Rev.*, **97**, 471–489, [https://doi.org/10.1175/1520-0493\(1969\)097<0471:MOPCT>2.3.CO;2](https://doi.org/10.1175/1520-0493(1969)097<0471:MOPCT>2.3.CO;2).
- Slocum, C. J., M. N. Razin, J. A. Knaff, and J. P. Stow, 2022: Does ERA5 Mark a New Era for Resolving the Tropical Cyclone Environment? *J. Climate*, **35**, 7147–7164, <https://doi.org/10.1175/JCLI-D-22-0127.1>.
- Wilks, D. S., 2019: *Statistical methods in the atmospheric sciences*. 4th ed., Elsevier, 818 pp., <https://doi.org/10.1016/C2017-0-03921-6>.
- Zehr, R. M., 1992: Tropical cyclogenesis in the western North Pacific. NOAA Tech. Rep. NESDIS 61, 189 pp., https://repository.library.noaa.gov/view/noaa/13116/noaa_13116_DS1.pdf.
-

END OF DOCUMENT