Earth and Space Science

RESEARCH ARTICLE
10.1029/2025EA004406

Key Points:

- Ice Water Path (IWP) directly measures storm presence and intensity, forming the basis for the proposed storm tracking method
- A machine learning method forecasts IWP by merging satellite infrared and microwave satellite data
- Improved storm tracking methods unlock underused satellite data to advance our understanding of convection

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

F. J. Tapiador,
francisco.tapiador@uclm.es

Citation:

Leganés, L. J., Prasanth, S., Navarro, A., Turk, J., Lee, G., Haddad, Z., et al. (2026). Machine learning-based forecasting of ice water path for storm detection and tracking using geostationary infrared and passive microwave data. *Earth and Space Science*, *13*, e2025EA004406. https://doi.org/10.1029/2025EA004406

Received 9 APR 2025
Accepted 11 DEC 2025

Author Contributions:

Conceptualization: Ziad Haddad
Data curation: Joe Turk, Peter Marinescu
Funding acquisition: Ziad Haddad, Francisco J. Tapiador
Investigation: Livia J. Leganés
Methodology: Livia J. Leganés, Ziad Haddad
Project administration: Ziad Haddad, Francisco J. Tapiador
Resources: Sai Prasanth, Ziad Haddad, Peter Marinescu, Francisco J. Tapiador
Software: Livia J. Leganés, Joe Turk, Ziad Haddad

Machine Learning-Based Forecasting of Ice Water Path for Storm Detection and Tracking Using Geostationary Infrared and Passive Microwave Data

Livia J. Leganés[1], **Sai Prasanth**[2], **Andrés Navarro**[3], **Joe Turk**[2], **Gyuwon Lee**[4], **Ziad Haddad**[2], **Jennie Bukowski**[5], **Peter Marinescu**[5], **Itinderjot Singh**[5], **Susan C. van den Heever**[5], and **Francisco J. Tapiador**[1]

[1]Department of Environmental Sciences, Institute of Environmental Sciences (ICAM), Earth and Space Science (ESS) Research group, University of Castilla-La Mancha, Toledo, Spain, [2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, [3]Universidad de León (ULE), Atmospheric Physics Group (GFA), Environmental Institute (IMA), León, Spain, [4]Department of Astronomy and Atmospheric Sciences, Center for Atmospheric REmote sensing (CARE), Kyungpook National University, Daegu, Republic of Korea, [5]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA

Abstract This study presents a proof of concept for a storm Detection, Delineation, and Tracking (DDT) method based on the vertical integral of condensed water and ice content—namely the Ice Water Path (IWP). Traditional DDT methods rely solely on regularly sampled variables, such as brightness temperatures, which do not directly capture storm evolution. In contrast, IWP is a direct indicator of storm presence and intensity, being zero when absent and high in convective cores. A machine learning approach is proposed to forecast IWP during gaps in passive microwave (PMW) observations. It integrates an initial IWP retrieval with the temporal evolution of geostationary IR data. The method was first evaluated using numerical weather model (NWP) simulations of convective episodes. Then, it was applied to 26 cases of near-coincident (20 min apart) PMW and IR observations of real storms. The prediction of IWP is addressed at three levels: (a) a cloud/no-cloud classification achieving probabilities of detection (POD) consistently above 0.60 and often exceeding 0.80; (b) a regression model with some challenges on the spatial structure forecast but normally achieving correlations above 0.60 (c) an intensity-based, three-class classification that outperformed simple variable persistence and demonstrated strong spatial coherence.

Plain Language Summary A new method for tracking storms is introduced, combining data from two types of satellite observations: geostationary infrared (IR) temperatures and passive microwave (PMW) estimates of the Ice Water Path (IWP). IR data provide frequent updates but only capture cloud-top temperatures, while IWP directly reflects the total amount of water and ice in a storm, making it a more reliable indicator of storm development. However, PMW observations are less frequent, creating gaps in the data. To address this, a machine learning approach is used to estimate IWP when direct PMW measurements are unavailable. The method starts with an initial IWP estimate, along with IR temperatures at that moment and at the next available time. Using this information, it predicts IWP at the later time step. The technique was evaluated at three levels: (a) identifying whether clouds are present or not; (b) predicting the storm's full structure and (c) a classification of storm intensity. The method was first tested using computer simulations of storms and then applied to real satellite observations. Results show that this technique effectively detects storms, predicts their structure, and classifies their intensity. These findings highlight the potential of IWP-based methods to improve storm monitoring and forecasting.

1. Introduction

Over the past two decades, an increasing number of spaceborne atmospheric remote-sensing instruments have generated a wealth of observations of convective storms—from isolated thunderstorms (Das et al., 2020; Jacobson et al., 2011; Rakesh et al., 2024; Zhang et al., 2016) and mesoscale convective systems (Moraes et al., 2020; Wang et al., 2022, 2024) to tropical cyclones (Li et al., 2023; May et al., 2024; Wang et al., 2021; Zheng et al., 2019; Zhuge et al., 2015). However, the full potential of these observations has yet to be realized in the study of convection due to the current limitations on Detection, Delineation, and Tracking (DDT) methods. Existing techniques (Fiolleau & Roca, 2013; Heikenfeld et al., 2019; Vila et al., 2008) primarily rely on

Supervision: Sai Prasanth, Andrés Navarro, Gyuwon Lee, Ziad Haddad, Francisco J. Tapiador
Validation: Sai Prasanth, Ziad Haddad, Francisco J. Tapiador
Writing – original draft: Livia J. Leganés
Writing – review & editing: Livia J. Leganés, Sai Prasanth, Andrés Navarro, Gyuwon Lee, Ziad Haddad, Jennie Bukowski, Peter Marinescu, Itinderjot Singh, Susan C. van den Heever, Francisco J. Tapiador

thresholding a single scalar variable and require spatial and temporal regularity. Geostationary IR observations are commonly used for their continuous coverage, but the brightness temperatures (TB) primarily reflect the cloud top, which does not adequately capture the full convective structure. To address this limitation, in this work, we explore the use of a more physically relevant variable, the Ice Water Path (IWP).

IWP is defined as the vertical integral of ice content within each air column (He & Liu, 2023; Mastro et al., 2022; Wang et al., 2014). It is zero in regions where no clouds are present; thus, storm genesis can be identified when this variable exceeds a defined threshold over a contiguous area of minimum spatial size. As convection intensifies, the area where IWP surpasses this threshold expands, and it contracts as the storm dissipates. If IWP data were available on a regular temporal and spatial grid, existing DDT methods could be applied. However, IWP retrievals rely on passive microwave radiometers (Ito & Masunaga, 2022; Kaur et al., 2022; Seto et al., 2023), which form a heterogeneous constellation and have a highly irregular coverage. This results in infrequent revisits to any given surface point, limiting the ability to continuously track a storm.

Given the complementary strengths of PMW and IR observations, combining both offers a promising approach. Geostationary IR observations provide high temporal resolution but are primarily sensitive to cloud-top radiances, lacking information on the deeper cloud structure. In contrast, PMW sensors are more directly related to the total ice content, as the longer wavelengths can penetrate through cloud layers, but they have coarser resolution and limited temporal sampling.The fundamental idea here is to rely on PMW sensors for their IWP retrievals while leveraging the continuous and high-resolution temporal evolution captured by geo-IR data (Tapiador, Kidd, Levizzani, & Marzano, 2004). This has been effectively utilized in global precipitation measurement projects, such as CMORPH (Climate Prediction Center morphing method, Joyce et al., 2004), IMERG (NASA's Integrated Multi-SatellitE Retrievals for the GPM mission, Huffman et al., 2014), and GSMaP (JAXA's Global Satellite Mapping of Precipitation, Aonashi et al., 2009) where accurate precipitation estimates from PMW sensors are propagated forward (or backward) in time using IR data. Other efforts have also applied machine learning techniques to enhance IWP retrievals from IR observations as in Tapiador, Kidd, Hsu, and Marzano (2004) and Tapiador et al. (200b) for PMW. More recently, Amell et al. (2022) used quantile regression neural networks with Meteosat-9 data to improve IWP retrievals and demonstrated the potential of ML within this context.

Following a similar rationale of the aforementioned products and studies, we propose a machine learning approach that avoids deriving IWP directly from IR data. Instead, it combines a reliable IWP retrieval from PMW at an initial time (t_0) with the changes in the spatial patterns of geo-IR—such as the shifting of local minima— to forecast the IWP at a subsequent time (t_1). Three algorithmic variants are explored: The first is a binary classification model that determines the presence or absence of clouds; the second is a regression approach aimed at predicting continuous IWP values; and the third, a three-class classification model that categorizes IWP into zero/low, moderate, and high regimes.

This study serves as proof of concept, aiming at demonstrating both the feasibility and scalability of the proposed methodology. Our initial tests are performed using simulations from Numerical Weather Prediction (NWP) models of convective events over the tropics. These simulations provide "perfect" data, meaning high spatial and temporal resolution free from instrumental errors, through which to perform controlled experiments, sensitivity tests, and to refine the algorithm under idealized conditions (Bolot et al., 2023; Gómez et al., 2020; Kukurugya et al., 2011; Leutwyler et al., 2017; Müller & Parlow, 2013; Oertel et al., 2019; Raut et al., 2021).

After to the simulation phase, the methodology is applied to real-world satellite observations of convective storms. For this purpose, a data set from the summer of 2023 was assembled over the tropics. Although PMW coverage is inherently sparse, a subdomain approach allows the extraction of hundreds of input–output pairs from each case. The results obtained, even with a limited data set, served to validate the conceptual framework.

2. Data

2.1. Simulations of Convective Episodes: Parametrization of IWP and IR Measurement

This study uses output from nine case study simulations of convection throughout the tropics and subtropics. These simulations have been conducted in support of the upcoming NASA INvestigation of Convective UpdraftS (INCUS; Dolan et al., 2023; Marinescu et al., 2024; Singh et al., 2025; van den Heever, 2021) mission. The simulations use the Regional Atmospheric Modeling System (RAMS; Pielke et al., 1992; Saleeby & van den Heever, 2013) and ERA5 reanalysis (Hersbach et al., 2020) as initial and lateral boundary conditions. We use data

Table 1
Name, Duration and Geographical Region of Each Simulation

Code	Duration (min.)	Location
ARG1.1-R	65	Lat: [28.21, 35.31] S Lon: [59.29, 67.90] W
ARG2.1-R	60	Lat: [28.21, 35.31] S Lon: [59.29, 67.90] W
AUS1.1-R	65	Lat: [4.98 N, 9.98 S] Lon: [137.95, 122.44] E
BRA1.1-R	60	Lat: [2.79 N, 11.02 S] Lon: [47.92, 68.07] W
DRC1.1-R	60	Lat: [2.79N, 11.02 S] Lon: [27.04, 16.15] E
PHI1.1-R	50	Lat: [20.29, 13.06] N Lon: [123.18, 114.313] E
PHI2.1-R	45	Lat: [20.09, 12.85] N Lon: [134.71, 124.48] E
USA1.1-R	70	Lat: [32.59, 25.319] N Lon: [89.18, 99.81] W
WPO1.1-R	70	Lat: [17.01, 9.94] N Lon: [143.07, 133.92] E

from the outermost domains of these simulations, which cover ~1,000 km v 1,000 km regions and have a horizontal grid spacing of 1.6 km, 232 vertical levels (with a maximum vertical spacing of 125 m), and a data output temporal resolution of 5 min Table 1 lists model output duration and geographical extent of the simulation domains.

These simulations provided an idealized environment for developing the methodology. However, since the model output does not provide in-line calculations of satellite observables, the IWP and IR brightness temperatures needed to be calculated or parameterized using the available variables. Additionally, to align better with the resolution of satellite data, the model output was coarsened—16 km for IWP and 4 km for IR.

The IWP was calculated as the integral of the total mixing ratio (Q_{ice}) of all frozen hydrometeors in the air column, following Equation 1.

$$\text{IWP} = \int_{\text{ground}}^{\text{top}} Q_{\text{ice}}\, dz \tag{1}$$

For the IR measurements we developed a Synthetic IR (SIR) based on two temperature indicators: one representing the high end of the cloud and the other marking the bottom of the top cloud layer (Equation 2):

$$\text{SIR} = \frac{1}{2}\left(T_{h_Q} + T_{h_{\text{IWP}}}\right) \tag{2}$$

The high end of the cloud (denoted as h_Q) is defined as the vertical level where Q_{ice} first exceeds 0.05 kg m^{-3}. It is determined by scanning downward from the highest level until this threshold is met. Similarly, the bottom of the top cloud layer (h_{IWP}) is identified using an IWP threshold of 0.015 kg m^{-2}. The thresholds chosen are consistent with temperatures below 235 K at altitudes of 10 km or higher, which corresponds to the typical temperature range for the cold upper layers of convective clouds. This IR temperature modeling is grounded in physical consistency, aligning with the fundamental definition of IR brightness temperature as a measure of cloud-top emission.

2.2. Construction of the Data Set With Polar-PMW and Geo-IR Satellite Measurements of Convective Storms

The data set was constructed using four satellite measurements over the same region: IWP at t_0 (retrieved from a first PMW pass), IR brightness temperatures at t_0 and t_1 (from geostationary IR), and the target output, IWP at t_1 (from a second PMW pass).

The primary challenge was obtaining a sufficient number of input–output pairs for training, given the sparse coincidence of PMW passes over the same region. Thus, in order to increase the number of usable cases, PMW data were gathered from multiple instruments, including the Microwave Humidity Sounder (MHS) aboard the METOP satellite series (A/B/C) and NOAA-18 and NOAA-19, the Advanced Technology Microwave Sounder (ATMS) aboard the NPP satellite, and the GPM Microwave Imager (GMI). Once the overlapping area between passes was localized, the measurements were remapped to a common 0.1° grid using cubic interpolation with the griddata function from SciPy (Virtanen et al., 2020).

IR measurements at t_0 and t_1 were obtained from the High Rate SEVIRI Level 1.5 Image Data at 0° on the Meteosat Second Generation (MSG) satellite. Among the available channels, IR_108 (10.8 μm) was selected due to its continuous 24-hr availability. While visible (VIS) channels, such as the High-Resolution Visible (HRV) band, could provide additional information on cloud microphysical properties, their dependence on daylight limits their applicability. Tests using additional IR channels did not yield significant improvements, leading to the adoption of a single-channel approach for simplicity. The brightness temperatures were interpolated onto a uniform 0.01° latitude-longitude grid.

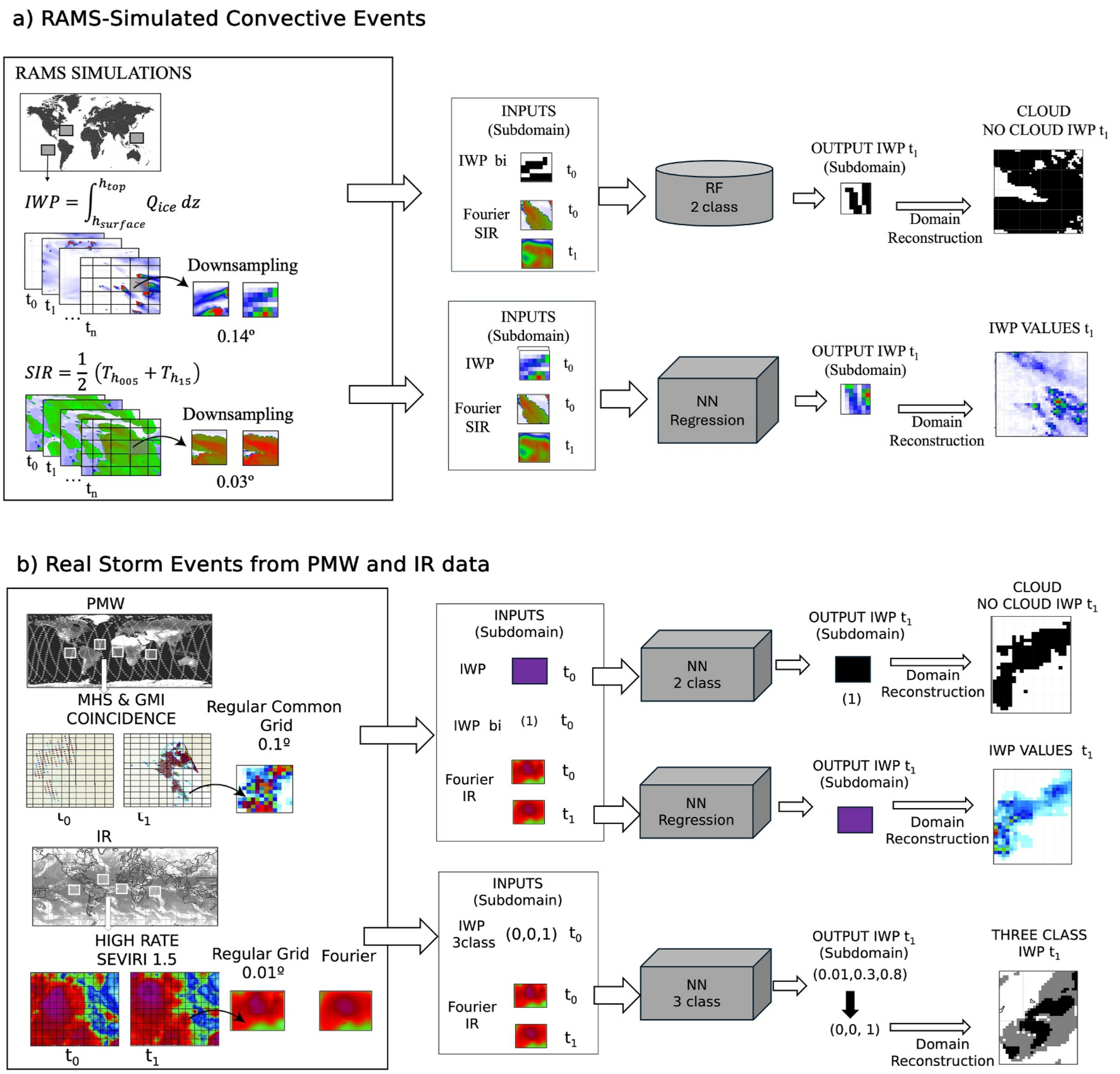

Figure 1. Schematics of the methodology for (a) model simulations; (b) satellite data.

The search for real-case scenarios was conducted in the tropical region during summer 2023, restricting satellite coincidences to a maximum 20-min window—as longer intervals may result in a loss of convective coherence. The variety of PMW sensors included in the analysis yielded 81 potential coincidences based on the defined criteria, specifically cloudy scenes within the geographical area during summer. Cases lacking storms or with minimal convection were excluded, along with those where small satellite swaths overlapped regions or where IR measurements primarily captured background noise. This resulted in 26 suitable cases.

3. Methods

Figure 1 illustrates the methodology step by step for both (a) application to RAMS-simulated events and (b) application to satellite measurements. The input for all models includes the IWP at the initial time (t_0) and IR temperatures at both t_0 and t_1. Depending on the algorithm, the IWP is used in its full range of values, discretized, or both. The predictions are validated by comparing them to the actual IWP values at t_1.

For Figure 1a, the evaluation on NWP simulations, inputs for the machine learning algorithm are derived from model parameters, where Q_{ice} represents the ice mass per cubic meter in the air column and SIR refers to Synthetic Infrared, a proxy for brightness temperature. SIR is calculated using the top and base temperatures of the first cloud layer and reduced via Fourier series. The variables are downsampled into subdomains: 7 × 7 grid for IWP and 28 × 28 for SIR. For binary classification, IWP is binarized (noted as bi) using a threshold of 0.015 $\text{kg}\cdot\text{m}^{-2}$, predicting whether each pixel will be cloudy or clear in the next time lapse. For regression, IWP values are used without binarization. Both predictions are made on a subdomain bases, thus a final step of reconstruction for the domain was done.

For Figure 1b, the application on real storm events from polar-PMW and geo-IR TBs, IWP data is taken from MHS and GMI retrievals, while IR data comes from SEVIRI. The area of overlap is interpolated onto a regular grid (0.1° for IWP, 0.01° for IR). Subdomains are formed: 1 pixel for IWP and 10 × 10 for IR. The same inputs are used for binary classification and regression, but with different neural network architectures. For the three-class classification, IWP at t_0 is categorized into no cloud—low IWP, low to moderate IWP and moderate to high IWP values. The neural network predicts class probabilities, which are post-processed to assign each pixel to a single class. The domains are also reconstructed as part of the post-processing.

3.1. Data Preprocessing

Both the simulation domains and the regions of overlap between satellite passes are quite large. This significantly increases the computational load when processed at once by a ML algorithm. At the resolution used in this study, a 10° × 10° latitude-longitude area contains thousands of inputs; thus, if it were to be directly fed to the ML model, it would result in an unnecessarily complex architecture. Instead, the IWP forecasting is carried out on smaller subdomains. Sensitivity tests showed that the optimal subdomain size differs between model data and satellite observations. For model data, a 7 × 7 grid for IWP and a 28 × 28 grid for IR were optimal. For satellite observations, a single IWP pixel paired with a 10 × 10 IR grid yields the best performance.

Further reduction of the number of inputs was achieved by applying a three-wave Fourier approximation to the brightness temperatures. The radial coefficients were then expanded using a fifth-order Taylor series (Equation 3). This approach effectively preserved the main features of the IR field while reducing the number of brightness temperature inputs from 784 to 42 for simulation data and from 100 to 42 for satellite observations. Only IR was approximated due to its higher resolution and role in capturing pattern changes, where the main components are sufficient. In contrast, the IWP was retained in full as the primary variable to maximize its contribution to the machine learning algorithms.

$$\mathrm{TB}(x,y) = a_0(r) + a_1(r)\cos(\theta) + a_2(r)\sin(\theta) + a_3(r)\cos(2\theta) + a_4(r)\sin(2\theta) + a_5(r)\cos(3\theta) + a_6(r)\sin(3\theta)$$

$$\text{with } a_i = \sum_{j=0}^{5} a_{ij} r^j \tag{3}$$

Lastly, the input–output pairs were normalized. IWP values were strictly positive and exhibit a right-skewed distribution, whereas IR Fourier Coefficients (FC) were both positive and negative with a narrower range. Although previous studies have shown that IWP can be effectively normalized using a log-transformation to approximate a lognormal distribution (e.g., Pfreundschuh et al., 2022), our tests indicated that when combining both IWP and IR features as inputs, the most robust and stable results were obtained by standardizing IWP values and rescaling IR coefficients between −1 and 1. This normalization process was applied independently to the two IWP estimates (t_0 and t_1) to avoid introducing prior knowledge into the learning process, ensuring that the model did not implicitly encode information about the boundaries of the future values.

3.2. Training Schemes

Instead of using a classical train–test split, we adopted a leave-one-out cross-validation approach. In Training Scheme 1 (TS1), the model is trained on all cases except the one being predicted. For instance, when predicting the third simulation, the model was trained on simulations 1, 2, and 4 through 9. Similarly, for satellite data, the model predicting case E was trained using all cases except E (i.e., cases A, B, C, D, and F through Z). At all times, predictions were made solely on truly unseen data.

Based on initial findings, we implemented an alternative procedure, Training Scheme 2 (TS2), which divides the data set into two categories: intensifying and weakening storms. The purpose of this scheme was to explore whether grouping the data according to distinct dynamical stages of convective evolution could improve the model's predictive performance. While TS2 uses a more specialized data set, this categorization could be automated in operational applications using ancillary IR-based metrics (e.g., fractional area of cold IR pixels or scene-averaged brightness temperature trends). Therefore, TS2 should be viewed as a methodological experiment designed to assess the added value of dynamically informed data partitioning.

Looking ahead to future operational settings, as the data set grows and more cases become available, it may be possible to develop a single, unified model for all storms, or separate models for each storm class (intensifying, weakening, and extreme events such as tropical cyclones), depending on the data set used for training.

It is worth noting that the training and prediction subsets were never intertwined. In addition, the evaluation of the algorithm on RAMS simulations and the real-world satellite measurement application was conducted separately, with no overlap between the two steps of the proof of concept.

3.3. Description of the Binary Classification Algorithm

The IWP was discretized, assigning a value of 0 for cloud absence and 1 for cloud presence. The binarization process differed between NWP simulations and satellite observations. For simulations, a fixed threshold of 0.015 kg m^{-2} was used to separate the two classes. However, satellite cases exhibited greater heterogeneity, introducing scalability challenges. To address this variability, a dynamic thresholding approach was implemented: for each predicted case, the threshold was set as 1% of the maximum IWP value at time t_0, allowing the data set to be scaled according to the specific characteristics of the case. Only the initial time was used, as incorporating IWP values from t_1 would introduce prior knowledge of storm evolution. The threshold, which varied from 0.01 to 0.16 kg m^{-2}, had an average value of 0.02 kg m^{-2}.

For the RAMS simulations data, a Random Forest (RF) classifier (Ho, 1995) was selected. RF is an ensemble learning method (Breiman, 2001) that constructs multiple decision trees, each trained on a random subset of the data. The final classification is determined by averaging the probabilistic outputs of all trees. The algorithm was implemented using the Scikit-learn library in Python (Pedregosa et al., 2011), with an architecture consisting of 50 decision trees evaluated using the Gini impurity criterion. The Gini index measures the impurity of a node, reaching zero when all samples belong to a single class. No maximum depth was set for the trees, allowing them to expand as necessary. The model inputs included the binarized IWP values at t_0 over a 7×7 grid, the IR FC at t_0 and t_1 (42 coefficients each), and the output was the binarized IWP at t_1 for the same 7×7 subdomain. Predictions were then compared to the actual binarized IWP at t_1.

The RF approach, however, was insufficient for satellite data, leading to the adoption of a Multilayer Perceptron (MLP). The chosen architecture consisted of an input layer with 86 neurons, three hidden layers with 90, 30, and 10 neurons each, and a single output neuron. Experiments showed that using a combination of activation functions led to better performance by improving learning stability. We used the ADAM optimizer and the mean absolute error (MAE) as the loss function. All neural networks were implemented using the Keras API (Chollet et al., 2015), with TensorFlow (Abadi et al., 2015) as the computational backend. The model architecture was kept minimal to encourage learning the underlying relationships between variables rather than memorizing specific cases. Additionally, an extra input—the non-binarized IWP value at t_0—was introduced to provide context on the initial state of the storm, which improved classification accuracy. The details of the MLP algorithm and training parameters are presented on Table 2.

The classification performance was evaluated using four indices (Cheng et al., 1993) computed from four outcomes: correctly predicted cloud pixels (CC), correctly predicted non-cloud pixels (nCnC), missed cloud pixels (CnC), and false positives (nCC). Probability of Detection (POD; Equation 4) measures the fraction of actual cloud pixels correctly identified by the model, with higher values indicating better detection. Conversely, the False Alarm Ratio (FAR; Equation 5) quantifies the proportion of non-cloud pixels incorrectly classified as clouds, with lower values being preferable. The Loss Function Factor (LFF; Equation 6) integrates both false positives and false negatives into a single metric, while the Skill Score (SS; Equation 7) compares the model's performance to that expected by random chance—where higher scores denote superior performance to just random guessing.

$$\mathrm{POD} = \frac{\mathrm{CC}}{\mathrm{CC} + \mathrm{nCC}}, \tag{4}$$

Table 2
Multilayer Perceptron Specifications for the Binary Classification of Real Storm Events

Binary classification MLP algorithm								
Architecture								
		Activation function						
			TS 2		Traning parameters			
Layer type	# Neurons	TS 1	Intensifying	Weakening	# Batch	# Epochs	Optimizer	Loss function
Input layer	86	-	-	-	128	8	ADAM	MAE
Hidden layers	90	LeakyReLu	LeakyReLu	PReLu				
	30	LeakyReLu	GELU	LeakyReLu				
	10	LeakyReLu	GELU	LeakyReLu				
Output layer	1	LeakyReLu	GELU	PReLu				

Note. TS1 and TS2 correspond to Training Scheme 1 and 2. The 86 input neurons correspond to the 42 × 2 Fourier Coefficients of the IR temperatures at t_0 and t_1, as well as the complete and binarized IWP at t_0.

$$\mathrm{FAR} = \frac{\mathrm{CnC}}{\mathrm{CnC} + \mathrm{CC}}, \tag{5}$$

$$\mathrm{LFF} = \frac{\mathrm{CnC}}{\mathrm{CnC} + \mathrm{CC}} + \frac{\mathrm{nCC}}{\mathrm{nCC} + \mathrm{nCnC}}, \tag{6}$$

$$\mathrm{SS} = \frac{\mathrm{CC} + \mathrm{nCnC} - k'}{\mathrm{CC} + \mathrm{nCC} + \mathrm{CnC} + \mathrm{nCnC} - k'}$$

$$\text{where } k' = \frac{(\mathrm{CC} + \mathrm{nCC})\,(\mathrm{CC} + \mathrm{CnC}) + (\mathrm{CnC} + \mathrm{nCnC})\,(\mathrm{nCC} + \mathrm{nCnC})}{\mathrm{CC} + \mathrm{nCC} + \mathrm{CnC} + \mathrm{nCnC}} \tag{7}$$

3.4. Description of the Regression Algorithm

For simulation data, the chosen MLP architecture consisted of 133 input nodes, which included 7 × 7 IWP values at t_0, along with 42 × 2 IR coefficients at both t_0 and t_1. The network featured two hidden layers with 40 and 35 neurons, and an output layer of 49 neurons—one for each IWP subdomain value at t_1. The model employed the Adam optimizer, LeakyReLU as the activation function, and Mean Squared Error (MSE) as the loss function.

When applying the methodology to satellite data, some modifications were necessary. The binarized IWP at t_0 was added as an additional input to highlight the importance of the IWP and to clarify whether a pixel contained clouds. To accommodate the increased complexity of the task, the number of neurons in the hidden layers was increased to 90 and 30, and a third hidden layer with 10 neurons was added. A combination of activation functions was also introduced to improve model performance. The optimizer and loss function remained unchanged. Detailed specifications of the algorithm can be found in Table 3.

In the application of this algorithm, an additional consideration was necessary. The regression was applied to the full range of continuous IWP values, without any discretization. The distribution of these values deviates from normality, as the highest IWP values are associated with the most developed convective columns and therefore represent only a small fraction of the domain. To achieve better representation of these extreme cases, which are important for the objectives of this study (i.e., locating and tracking tropical storms), a data augmentation method was applied. An example of this kind of methodology can be found in Troncoso et al. (2025). Here, it was applied as follows. Samples where IWP exceeded 3 kg m^{-2} were extracted and replicated three times in the training data set. This replication factor was empirically selected, and no synthetic cases were generated for this purpose.

3.5. Description of the Three-Class Classification Algorithm

This algorithm was specifically designed for real-world satellite cases; it serves as a midpoint between the previous two approaches. Ideally, the regression model would perform seamlessly on observational data.

Table 3
Multilayer Perceptron Specification for the Prediction of Ice Water Path as a Continuous Variable on Real Storm Events

Regression MLP algorithm

Architecture

			Traning parameters					
				Number of EPOCHS				
					TS 2			
Layer type	# Neurons	Activation function	# Batch	TS 1	Intensifying	Weakening	Optimizer	Loss function
Input layer	86	-	128	1	1	8	ADAM	MAE
Hidden layers	90	PReLu						
	30	LeakyReLu						
	10	GELU						
Output layer	1	LeakyReLu						

Note. The 86 input neurons correspond to the 42 × 2 Fourier Coefficients of the IR temperatures at t_0 and t_1, as well as the complete and binarized IWP at t_0.

However, in its absence, the next best approach to a functional DDT is an algorithm that determines whether a grid point contains clouds and its intensity.

In this case, the IWP discretization was done into three categories: no cloud or low IWP (class 0), low to moderate IWP (class 1), and moderate to high IWP (class 2). The threshold between classes 0 and 1 was consistent with the binary classification approach, set at 1% of the IWP at t_0. The boundary between classes 1 and 2 was fixed at $0.5\ \mathrm{kg \cdot m^{-2}}$.

The MLP model had an architecture with 86 input neurons, two hidden layers with 150 and 50 neurons, and an output layer with 3 neurons. A combination of activation functions was used: LeakyReLU for the hidden layers and SoftMax for the output layer. The loss function was Cross-entropy and the optimizer, ADAM. The model's inputs are the FC of the IR subdomain at t_0 and t_1, along with the IWP at t_0, which was represented as a 3-column array. Each column corresponded to one of the three IWP classes, where a value of 1 was assigned if the pixel belonged to that class and 0 otherwise.

The model's output was structured similarly, providing a probability distribution across the three classes. The final classification was determined based on the class with the highest probability. For example, an output of (0.3, 0.1, 0.6) would indicate a 60% probability of belonging to class 2. The algorithm specifications are detailed in Table 4.

Table 4
Multilayer Perceptron Specifications for the Classification of the Ice Water Path Into Three Classes on Real Storm Events

Three class classification MLP algorithm

Architecture

			Traning parameters			
Layer type	# Neurons	Activation function	# Batch	# Epochs	Optimizer	Loss function
Input layer	87	-	128	Early stopping	ADAM	Categorical cross-entropy
Hidden layers	150	LeakyReLu				
	50	LeakyReLu				
	30	LeakyReLu				
Output layer	3	Softmax				

Note. The 87 input neurons correspond to the 42 × 2 Fourier Coefficients of the IR temperatures at t_0 and t_1, as well as the IWP at t_0 in the three class format.

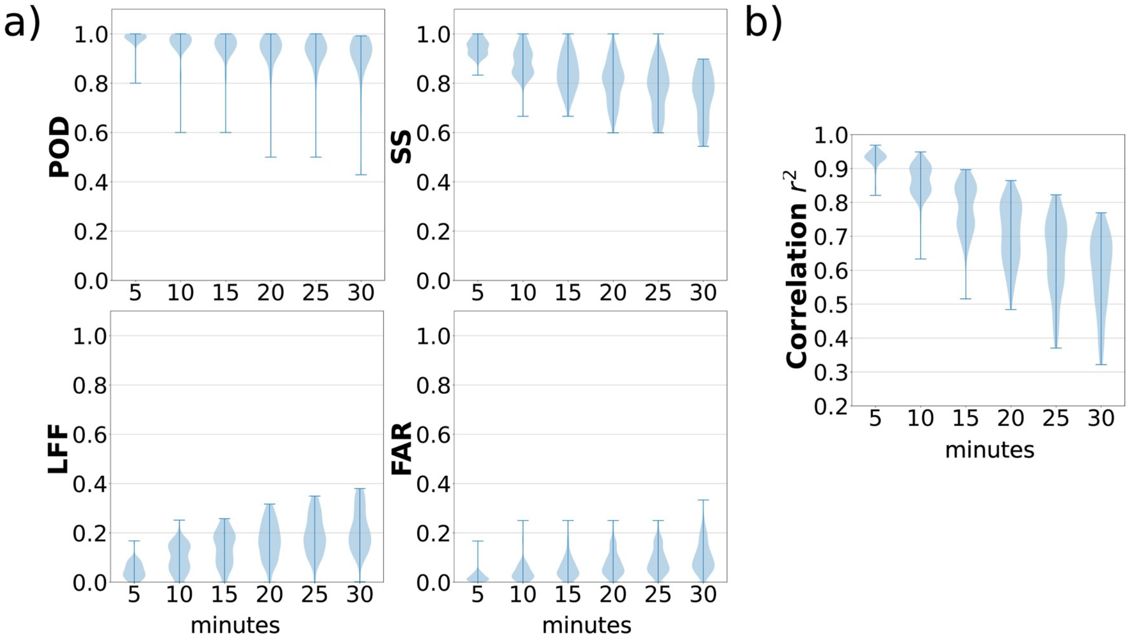

Figure 2. Violin plots of the (a) metrics for the Random Forest classification of cloud and no-cloud classes, including probabilities of detection, skill score, loss function factor, and false alarm ratio. (b) Pearson correlation between the multilayer perceptron forecast of continuous ice water path and the reference values. The *x*-axis represents the time difference (in minutes) between t_0 and t_1. The distributions are derived from all predictions made within each specified time interval.

4. Results

4.1. Analysis of Algorithm Efficacy in RAMS-Simulated Convective Events

We tested six forecasting intervals—5, 10, 15, 20, 25, and 30 min—between the initial state (t_0) and the prediction time (t_1). Unlike PMW satellites, which typically provide at most two observations per convective event, the simulations provided storm evolution data every 5 min. Longer intervals were deemed unnecessary since most convective cells normally lose coherence beyond 30 min.

Figure 2 shows the metrics for both binary classification and regression algorithms. These metrics are presented as violin plots, which are constructed from all the predictions made on simulations and sorted by the time interval between the initial time and the forecast/target time. The width of the plot represents the density of values at different levels, with the center of each distribution indicating the median value. The surrounding areas illustrate the spread and variability of the predictions.

In the binary classification task, the RF classifier maintained a probabilities of detection (POD) above 0.8 across all intervals, while the SS decreased with longer intervals but remained above 0.6. Error metrics, including the LFF and the FAR, increased over time but stayed below 0.4, indicating low rates of both false positives and false negatives. For the regression task, the Pearson correlation between the MLP's continuous IWP predictions and the nominal truth reached approximately 0.95 at a 5-min interval. However, the decline in correlation with increasing intervals was more pronounced than in the binary case. This is expected, as the regression task is inherently more complex—predicting continuous IWP values (ranging from 0 to 3 kg m^{-2}) from a non-normal distribution, where even small changes can be significant. Nonetheless, the MLP produced coherent predictions, with correlation peaks of around 0.8 and 0.6 even at a 20-min interval.

Based on these results, a maximum of 20-min forecasting interval was adopted for real data, balancing the need for sufficient case availability with the algorithm's ability to capture storm evolution. Shorter intervals, although potentially more accurate, would have drastically reduced the number of available PMW coincidences.

Figure 3 includes part of the evolution of AUS 1.1 simulation within a 20-min window for both the RF classification (top row) and the NN regression (bottom row). The first column displays the storm's initial state, while the second and third columns show the actual IWP evolution and the model's output, respectively. Multiple

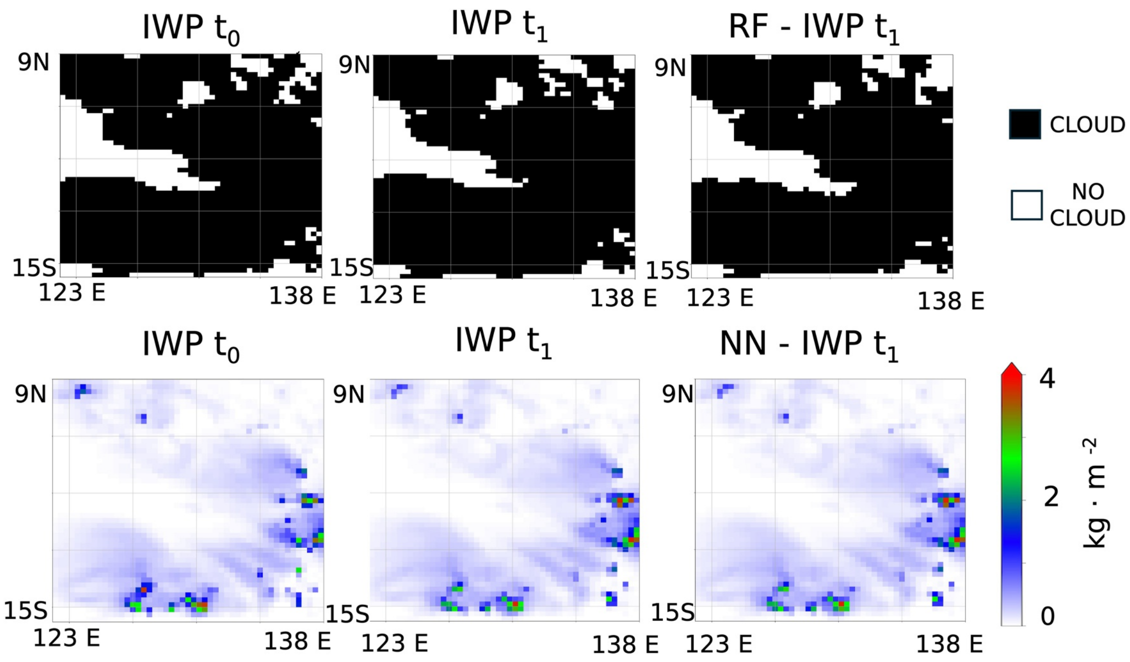

Figure 3. Ice water path (IWP) evolution and algorithms forecast for AUS1.1 with a 20-min interval between t_0 and t_1. The top row presents the binary classification of IWP at t_0, t_1, and the random forest prediction at t_1, where white represents no clouds (class 0) and black indicates clouds (class 1). The second row illustrates the same storm with IWP as a continuous variable at t_0 and t_1, with the final column showing the NN prediction of the storm's evolution at t_1.

convective cells were concentrated in the southern and eastern parts of the domain, with less developed clouds dominating in the north, the center is free from clouds at either time. The RF classifier accurately captured the overall cloud structure with some overestimation on the reduction in cloud coverage near the domain's center. In contrast, the NN prediction is able to predict subtle shifts in regions with higher IWP values, shown in red and green on the map, and is also able to predict the stability on structure of lower IWP regions.

The results in these idealized data conditions are promising, as they allowed for testing the core concept of the methodology. Since these conditions are free from real-world challenges, poor performance here would suggest that the methodology would likely not succeed with actual observational data. Thus, the next step is to assess the performance and applicability across the full range of variations found in real-world measurements.

4.2. Assessing Algorithm Accuracy Using PMW and IR Satellite Data From Real Storm Events

The performance of each algorithm was evaluated across 26 storm events, labeled from A to Z. Cases A to S correspond to intensifying storms, and cases T to Z to weakening storms. The metrics are discussed across all cases but, since discussing each one individually would be impractical, we focus on at least two representative examples for each algorithm. The goal was to include a broad range of examples; therefore, different cases were selected for each algorithm, with maps for the remaining cases provided in the Supplemental Information. The specific cases chosen were not critical, as the aim was to assess the general patterns of spatial coherence and the potential influence of different cloud configurations on algorithm performance.

In addition to binary classification and full IWP regression, a three-class classification approach is also discussed. While binary classification alone may be too simplistic for a useful DDT, the three-class algorithm, which captures both location and intensity, could provide sufficient accuracy.

4.2.1. Binary Classification Performance

Figure 4 presents the classification metrics (POD, SS, FAR, and LFF), organized into four rows. The first two correspond to TS1, where cases are not categorized, while the third and fourth rows correspond to TS2 (weakening vs. intensifying storms). With TS1, POD values were consistently high—typically above 0.6 and often reaching 0.8—alongside strong SS, indicating that the neural network effectively predicts cloudy grid points,

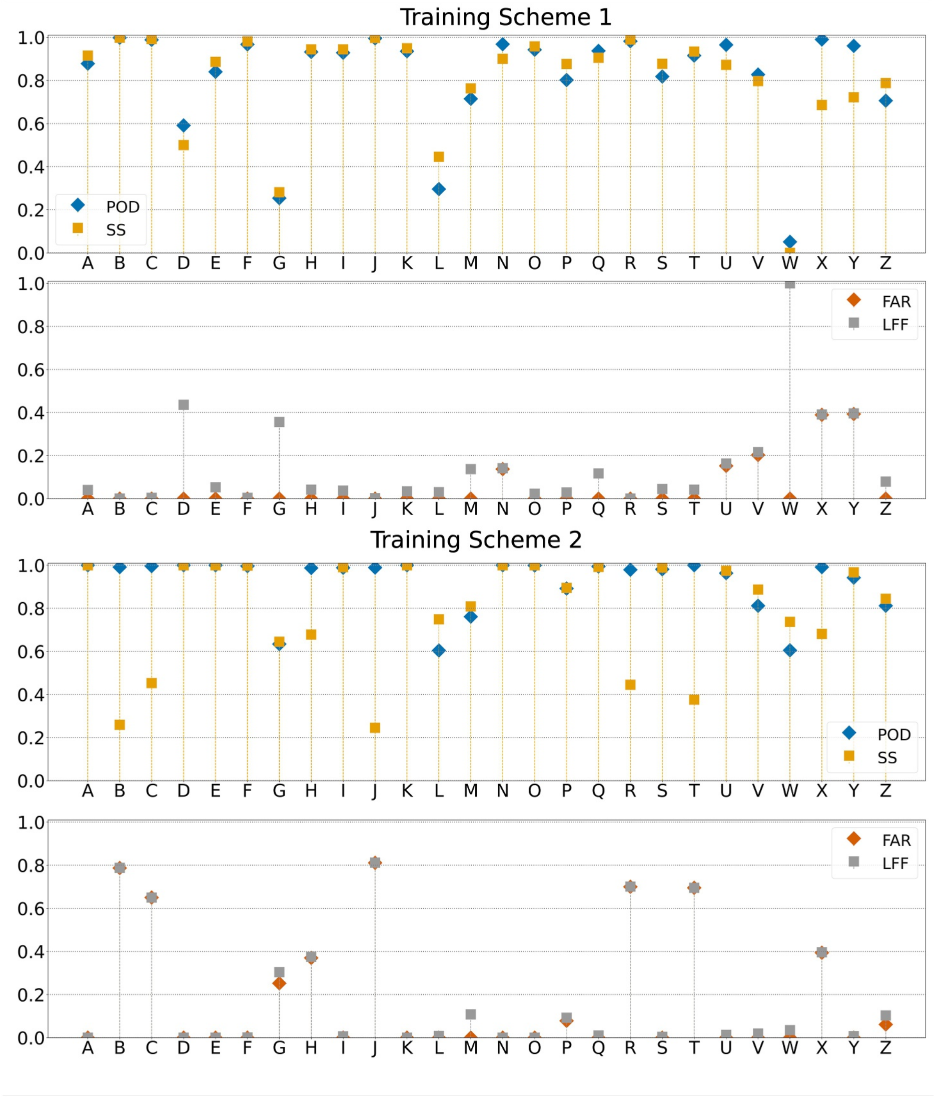

Figure 4. Binary classification metrics for the multilayer perceptron trained with Training Scheme 1 -first and second row—and Training Scheme 2 -third and fourth row-across the 26 convective episodes. The first and third rows present the success metrics: probabilities of detection (blue diamonds) and SS (orange squares), while the second and fourth rows display the misclassification metrics: FAR (dark red diamonds) and LFF (gray squares).

outperforming random guessing. However, four cases (D, G, L, and W) exhibited poorer performance, for case W the NN predicted the entire domain as cloudy (class 1). Under TS2, the POD improved across all cases, and although the SS showed more variability, it generally maintained higher values. This can be explained with the separation of intensifying and weakening storms in TS2 reduced contradictory examples in the data set, yielding a more robust detection overall.

At first glance, the combination of high POD values with elevated FAR and LFF may seem contradictory. However, high POD shows that the model is effectively detecting cloud pixels and high FAR and LFF indicate that some no-cloud pixels are being misclassified as clouds. This can be attributed to the discretization of a continuous physical variable, which naturally introduces a bias toward one of the classes. Additionally, the imbalance between cloud and no-cloud pixels means that, even though the algorithm detects most cloud events, the relatively small number of cloud pixels makes LFF particularly sensitive to false negatives.

Figure 5 presents two examples of the binary forecast: Cases A and D. The first column presents binarized IWP at t_0, the second shows binarized IWP at t_1, and the third and fourth columns display NN classification results using

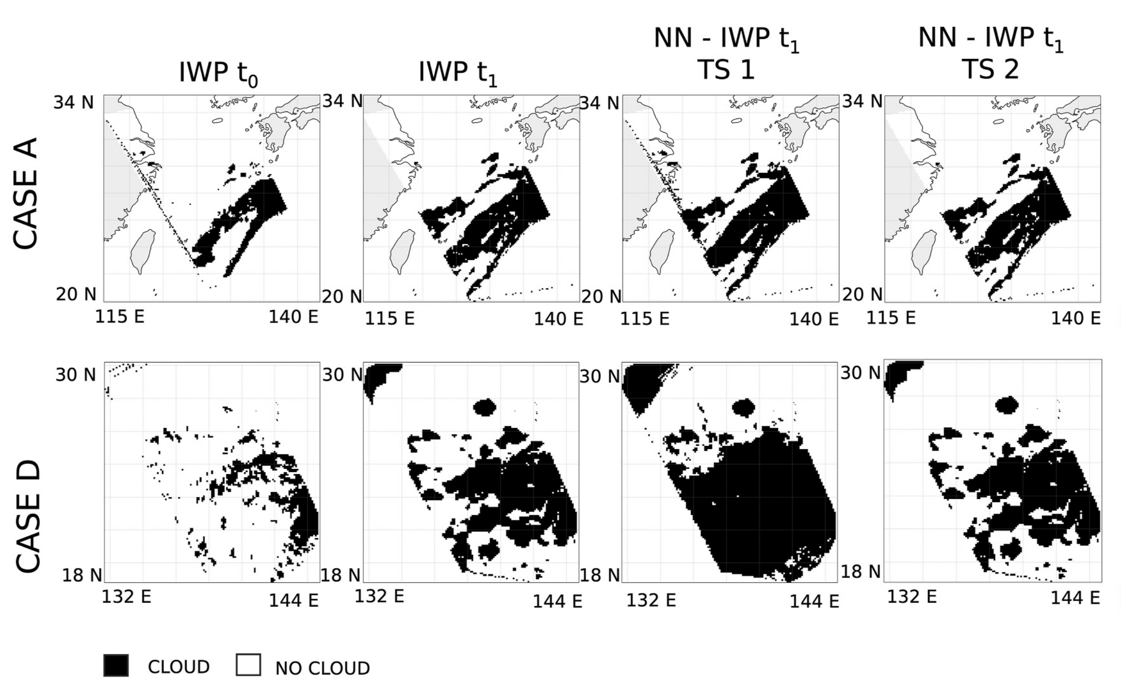

Figure 5. Example maps showing Ice Water Path (IWP) evolution and NN-based classification into two categories. The columns, from left to right, show the binary classification of IWP values at the initial time (t_0), at t_1 (20 min later), and the classified IWP values at t_1 using the two training schemes (Training Scheme 1 and Training Scheme 2). Black color for cloudy regions and white color for no-cloud regions. The top row corresponds to Case A, and the bottom row to Case D.

TS1 and TS2, respectively. In Case A, clouds were localized at the center of the domain in a clamp-like formation. By t_1, cloud coverage expanded, and new branches of the storm formed. Both TS1 and TS2 capture these changes, but TS2 more accurately delineates the evolving structure by better identifying the central no-cloud region. Although TS1 produces a coherent forecast, it appears to overemphasize artifacts from the western edge at t_0, which likely result from the edge of the satellite swath and are amplified by the interpolation technique.

For Case D, the change on IWP was more pronounced. At t_0, the clouds in the center of the domain were scattered, with additional cloud coverage in the northern part. By t_1, the smaller convection nodes increased in size and merged, leading to a more substantial cloud presence in the northern region. While TS1 could track some of the evolution of the northern cloud node, it lost coherence in the region from 24°N to the southern edge of the domain. However, TS2 was capable of accurately predicting the changes in the structure.

4.2.2. Regression Performance

The common regression metrics are shown in Figure 6: Pearson's correlation coefficient (r^2), MSE, MAE, and mean bias. The results are divided by training scheme, with the first two rows corresponding to TS1 and the bottom two rows to TS2.

For both training schemes, the mean bias was near zero for positive values but ranged from −0.1 to −0.2 for negative values, which may suggest an underprediction tendency. TS2 reduced this bias, especially in weakening cases. The MAE averaged 0.1 units (3.3% relative error based on a maximum IWP of 3 kg m^{-2}). Differences between MSE and MAE were generally small, suggesting no significant outliers, except in case M, where the MSE was 0.1 higher than the MAE, warranting further discussion. Case M posed a particular challenge for the NN. At t_0, convection was organized into small convective cells in the southeast, with no stratiform clouds but a completely cloud-free northwest region. After 20 min, most convective cells weakened, except for one that intensified significantly. Since this case did not fit into the intensifying or weakening categories, TS2 did not outperform TS1. Under TS1, the algorithm predicted a general increase in IWP values, maintaining smaller convective cells and introducing a broad region of weak IWP values. TS2, on the other hand, predicted

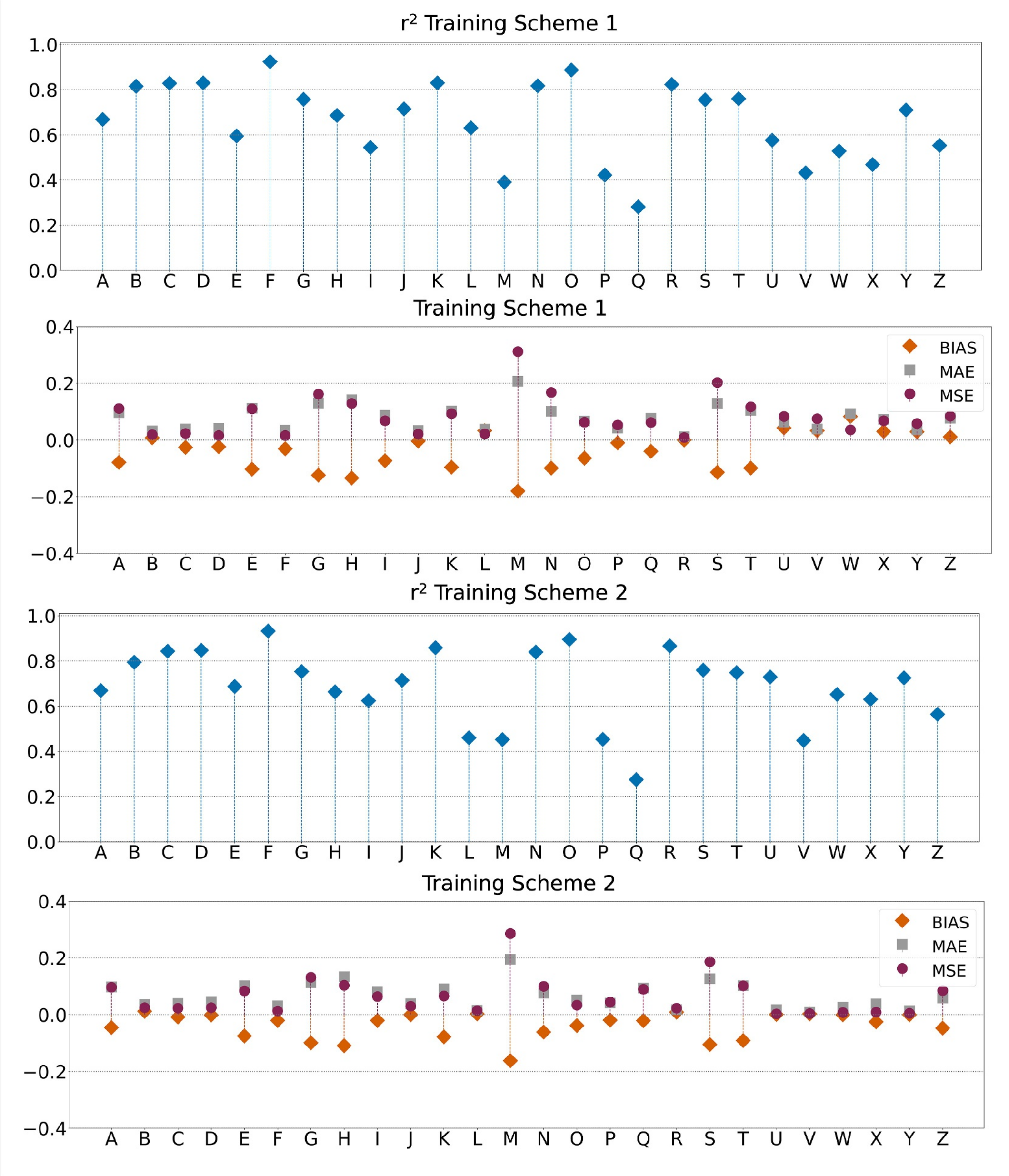

Figure 6. Regression metrics for the multilayer perceptron trained on Training Scheme 1 and Training Scheme 2 across the 26 convective episodes. The first and third rows present the Pearson correlation coefficient, while the second and fourth rows display the mean bias (orange squares), mean squared error (dark red circles), mean absolute error (gray squares).

intensification in weaker convective nodes but failed to forecast stratiform clouds and underestimated the intensification of the primary convective cell.

In terms of correlation, TS2 showed improvement for weakening cases. For intensifying cases, results were more mixed, with cases like E showing improvement and others like L showing a decline. However, most cases achieved r^2 values above 0.6, with many exceeding 0.8.

Figure 7 assesses the spatial coherence of this algorithm through cases U, H, and C. Each row presents the storm's evolution from t_0 (first column) to t_1 (second column, nominal truth), along with the NN predictions using TS1 (third column) and TS2 (fourth column). In Case U, the storm initially had moderate IWP values in the central-western part of the domain. After 20 min (t_1), the maxima dissipated, and the cloud region split into two weaker

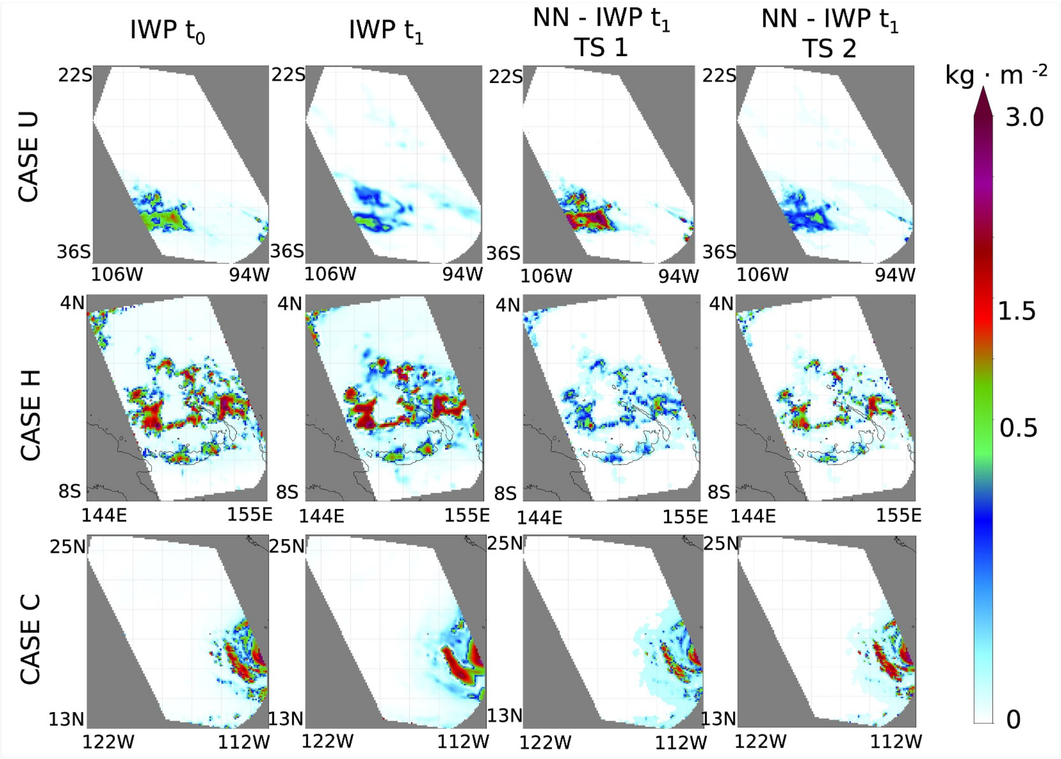

Figure 7. Example maps showing ice water path (IWP) evolution and NN forecast. The columns, from left to right, show IWP values at the initial time (t_0), at t_1 (20 min later), and the predicted IWP values at t_1 using the two training schemes (Training Scheme 1 and Training Scheme 2). The top row corresponds to Case U, second row to CASE H and the bottom row to Case S.

structures. TS1 predicted further intensification of the maximum convection and little change in storm shape, likely due to the underrepresentation of weakening cases in the global data set. TS2 captured the weakening nature of the storm but failed to predict the structural changes.

For Case H, the initial IWP field was characterized by scattered convective cores distributed across the domain. By the prediction time, the convection intensified uniformly while maintaining a stable shape. Both training schemes resulted on a NN model that incorrectly predicted the weakening trend; however, TS1 introduced a stronger negative bias compared to TS2, which is consistent with the metrics in Figure 6.

Case C represents a segment of Hurricane Hilary's evolution on 18 August 2023, at its peak intensity as a Category 4 hurricane. The initial scan at t_0 (09:02 UTC) showed minimal cloud coverage beyond the main structure. By t_1 (09:22 UTC), convective bands had intensified, forming a more compact structure with a low-IWP region around the core. TS1 partially captured the emergence of the low-IWP region but predicted weakening of the core and missed the restructuring of the bands. TS2, however, correctly identified both the low-IWP region and core intensification, though it only partially captured the changes in the band configuration.

The TS2 training scheme improved predictions by reducing bias and better capturing storm evolution. The application of the proposed methodology on storms with similar dynamics or in a similar part of the life cycle could further improve DDT techniques.

4.2.3. Three Class- Classification Performance

Figure 8 presents the metrics for this algorithm, starting with the confusion matrices for TS1 (left) and TS2 (right) in the first row. The confusion matrix evaluates the algorithm's performance by comparing predicted and true IWP values. Ideally, the matrix would be diagonal, indicating perfect predictions. Class 0 (no or low IWP), the most common category, was accurately classified by both training schemes, with the MLP correctly predicting all 219,176 pixels of this class. This means no cloud was incorrectly predicted where none existed. The algorithm also performed well for Class 1 (low to moderate IWP), with TS1 correctly classifying 84% of Class 1 pixels and

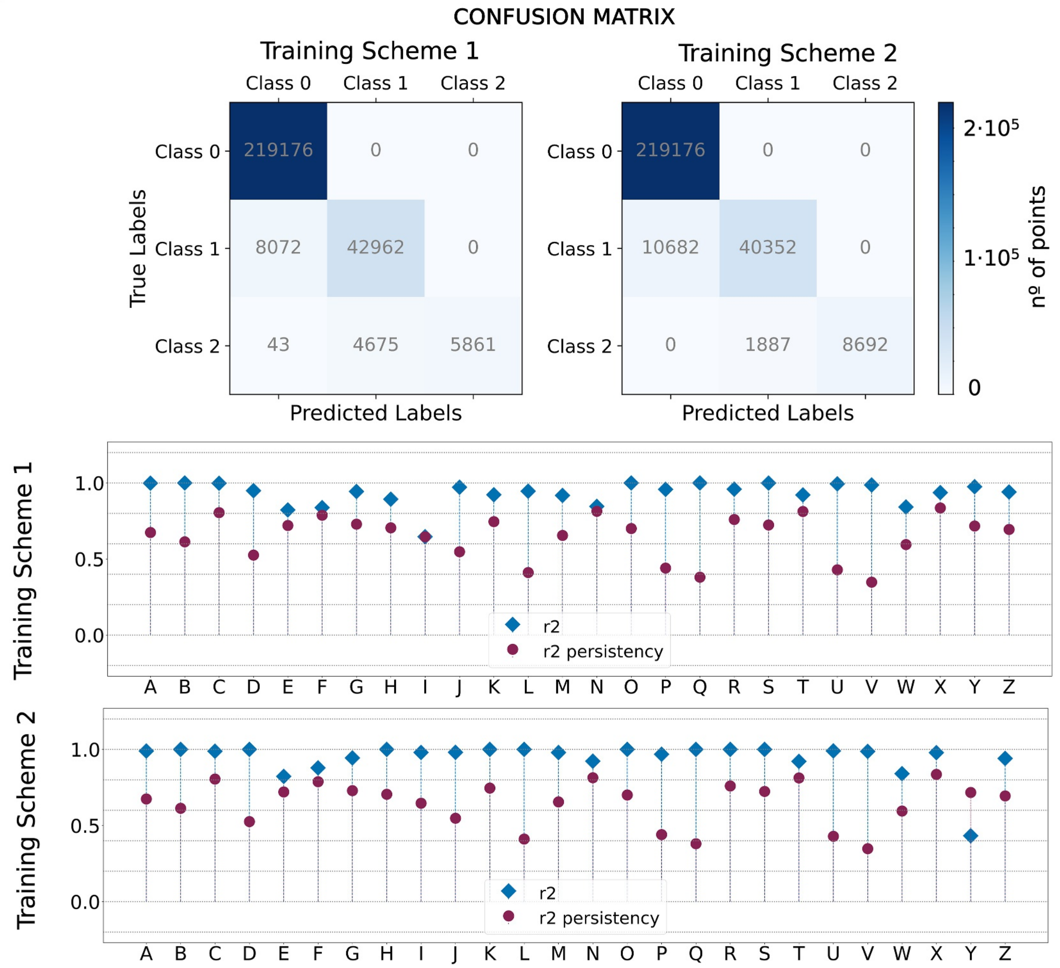

Figure 8. Metrics evaluating the MLP's performance in forecasting ice water path (IWP) across three classes, trained under two different schemes (Training Scheme 1 and Training Scheme 2). The top row shows confusion matrices, comparing true labels (IWP at t_1) with NN-predicted labels. The second and third rows display Pearson correlation coefficients: the correlation between NN-predicted IWP values at t_1 and true IWP values at t_1 (performance metric, blue diamonds), and the correlation between IWP values at t_0 and t_1 (persistence, dark red circles).

TS2 achieving 79% accuracy. Notably, no overestimation of storm intensity occurred, as no Class 1 pixels were misclassified as Class 2. TS2 outperformed TS1 for Class 2 (moderate to high IWP), correctly classifying 82% of the pixels without misclassifying any as Class 0, avoiding false alarms for intense convective cores. TS1 achieved only 55% accuracy, with 0.4% of Class 2 pixels misclassified as no cloud.

The second and third rows of Figure 8 show Pearson correlation coefficients for each training scheme. The first correlation (blue) evaluates how well the NN classifies IWP values at t_1 compared to the actual IWP values, serving as a direct measure of model performance.The second correlation (dark red) captures the persistence of IWP over time by measuring the relationship between IWP at t_0 and t_1. In other words, it indicates how much of the IWP state at t_1 can be explained solely by its previous value at t_0, meaning how much the variable have changed. By comparing these two correlations, we can determine whether the NN is truly learning new patterns or simply relying on persistence.

The performance correlations generally ranged between 0.8 and 1.0, with a minimal difference between the training schemes, except for case Y, where TS2 showed a notable decrease. In this case, the NN trained with TS2 only predicted Class 2 pixels but misclassified them as Class 1, missing a great part of cloud coverage. Overall, TS1 performed better for Class 1 but failed to classify Class 2 pixels correctly. Since Class 2 was less represented in the data set, errors in this category had less impact on the correlation coefficient.

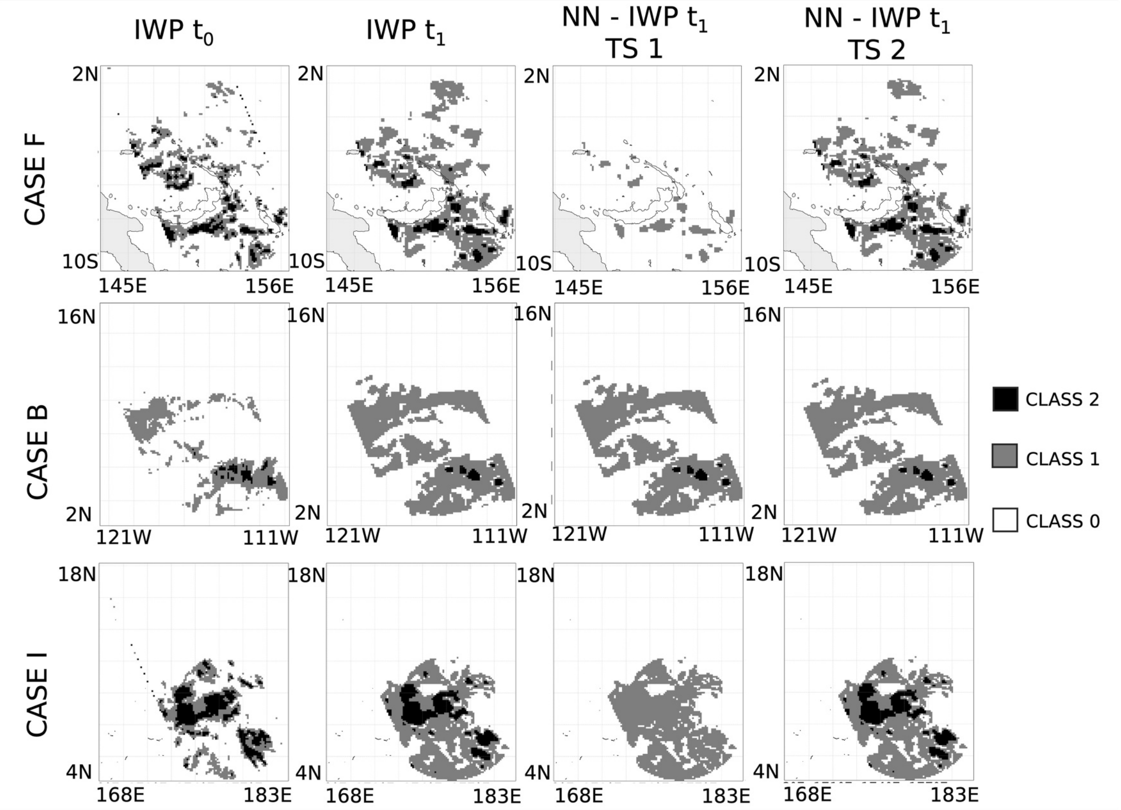

Figure 9. Example maps showing ice water path (IWP) evolution and NN classification in three classes. The columns, from left to right, show IWP values at the initial time (t_0), at t_1 (20 min later), and the classified IWP values at t_1 using the two training schemes. White color shows class 0 (no clouds), gray color class1 and black color class 2. The top row corresponds to Case F, second row to Case B and bottom row to Case I.

Regarding the comparison between the two correlations, the NN typically outperformed persistence. The persistence correlation was usually above 0.5, as some continuity was expected on the storms; however, there was no direct relationship between higher persistence and better ML classification. Regardless of the training scheme, the algorithm demonstrates the ability to identify and leverage fundamental physical relationships within the data, contributing beyond mere continuity or persistence of IWP values.

Figure 9 shows the algorithm's spatial coherence through cases F, B, and I. The columns display: (a) initial IWP values at t_0, (b) observed IWP at t_1 (reflecting changes over 20 min), and (c) NN predictions at t_1 using TS1 and TS2. Class 0 (low to moderate IWP values) are shown in white, Class 1 (moderate to high values) in gray and Class 2 (moderate to high) in black.

In case F, between t_0 and t_1, class 1 expanded and new cloudy regions appeared, while Class 2 transitioned from scattered points to more compact cores, becoming more concentrated in the southeastern part of the domain. TS2 closely matched the observed changes, successfully capturing the intensification and expansion of moderate and strong IWP areas. In contrast, TS1 misclassified Class 2 pixels as Class 1, resulting in a prediction that only captured the higher IWP values and overlooked the rest of the storm structure, similar to Case Y discussed earlier. For case B, Class 2 pixels decreased between t_0 and t_1 but reorganized into larger, more compact cores. Class 1 became more developed, filling most of the domain. Both TS1 and TS2 performed equally well.

The storm in case I underwent significant intensification. Initially, at t_0, the storm was compact, with Class 2 concentrated in the center of the domain. By t_1, the storm had intensified, with Class 2 spreading outward and Class 1 regions increasing in coverage. TS1 failed to classify Class 2 pixels correctly, labeling them as Class 1, although it captured the general cloud shape. TS2, on the other hand, accurately predicted the expansion and intensification of both Class 1 and Class 2.

5. Conclusions

This study illustrates that integrating IR temperatures and PMW estimates through ML algorithms allows to forecast the IWP, a key variable for DDT. The convective core of a storm is characterized by strong updrafts, deep cumulonimbus clouds, and the highest IWP values, while the stratiform region surrounds the core with more widespread, moderate precipitation and IWP values. The anvil or outflow region, located at the top of the system, consists of cirrus clouds with the lowest IWP values, indicating a diffuse distribution of ice. Indeed, correctly mapping the IWP within a convective system is tantamount to pinpointing the critical physical processes.

The method presented here opens new research avenues to elucidate the physics of convection. While missions such as NASA INCUS (van den Heever, 2021; Marinescu et al., 2024) will boost our understanding of the dynamics and microphysics of convective clouds thanks to increased temporal resolutions (30–90 s) measurements, large-area estimates of the IWP are expected to be instrumental to predict storm evolution in the Tropics. So far, our research can help to understand the dynamics of the convective mass flux on the tropics and how the environmental condition affects them. New data on microphysics would certainly improve performances.

The validation provides insights into what performance could be expected before INCUS data is integrated into a fully-fledged system. The current correlation between predicted and observed values was strong, with most values exceeding 0.60 and some nearing 0.80. The error rates were minimal, contained within the range of (−0.30, 0.30), and the absence of large outliers was reflected in the similarity between MAE and MSE. The model is inclined to slightly underpredict IWP values, a feature that would require more research and data to be palliated.

A regularly spaced IWP field could be used to forecast storm evolution and provide more accurate and precise warnings to populations in the trajectory of tropical storms.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The convective event simulations were generated using the RAMS model in the context of the INCUS mission. The RAMS model code (in the *src* directory) and the namelists (in the *run* directory) for the simulations used in this manuscript are available on GitHub (https://github.com/CSU-INCUS/RAMS/releases/tag/Database_V0.1).

Satellite data used in this study were selected based on criteria described in the manuscript (PMW passes within a 20-min window over the same region, located within the tropics during summer 2023). All PMW products can be accessed through NASA's GPM PPS-STORM data source: https://storm.pps.eosdis.nasa.gov/storm/data/Service.jsp?serviceName=Order. Researchers can retrieve the required PMW scenes by querying the PPS-STORM service using the event of interest date, time, and location. Geostationary infrared data are available from the EUMETSAT MSG HRSEVIRI repository, which provides a user-friendly interface for selecting data by date, time, and location: (https://data.eumetsat.int/data/map/EO:EUM:DAT:MSG:HRSEVIRI).

For researchers interested in applying this proof-of-concept approach to other data sets, the software may be adapted to specific data characteristics. Contact the corresponding author for assistance in tailoring the methodology to particular cases.

Acknowledgments
FJT acknowledges the financial support of the Spanish Ministerio de Ciencia e Innovacion and Agencia Estatal de Investigacion, MCIN/AEI/10.13039/501100011033, ERDF A way of making Europe and the European Union NextGeneration EU/PRTR (Projects: PID2022-138298OB-C22; TED2021-131526B-I00; PDC2022-133834-C21; PID2019108470RB-C21). FJT also acknowledges the support of the University of Castilla-La Mancha (2022-GRIN-34329, 2025-GRIN-38492). LJL also acknowledge the funding from the Spanish Ministerio de Ciencia e Innovación and Agencia Estatal de Investigación PRE2020-092986. GL acknowledges the funding of Korea Meteorological Administration Research and Development Program under Grant KMI2022-00310 and the support of the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2025-02242970). This research was partly conducted at Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. SP, ZSH, JB, PM, IS, and SH acknowledge the support from NASA INCUS. INCUS is a NASA Earth Venture Mission, funded by NASA's Science Mission Directorate and managed through the Earth System Science Pathfinder Program Office under contract number 80LARC22DA011.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software Retrieved from https://www.tensorflow.org

Amell, A., Eriksson, P., & Pfreundschuh, S. (2022). Ice water path retrievals from Meteosat-9 using quantile regression neural networks. *Atmospheric Measurement Techniques Discussions*, *2022*(19), 1–28. https://doi.org/10.5194/amt-15-5701-2022

Aonashi, K., Awaka, J., Hirose, M., Kozu, T., Kubota, T., Liu, G., et al. (2009). GSMaP passive, microwave precipitation retrieval algorithm: Algorithm description and validation. *Journal of the Meteorological Society of Japan*, *87A*, 119–136. https://doi.org/10.2151/jmsj.87A.119

Bolot, M., Harris, L. M., Cheng, K. Y., Merlis, T. M., Blossey, P. N., Bretherton, C. S., et al. (2023). Kilometer-scale global warming simulations and active sensors reveal changes in tropical deep convection. *npj Climate and Atmospheric Science*, *6*(1), 209. https://doi.org/10.1038/s41612-023-00525-w

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Cheng, M., Brown, R., & Collier, C. G. (1993). Delineation of precipitation areas using Meteosat infrared and visible data in the region of the United Kingdom. *Journal of Applied Meteorology*, *32*(5), 884–898. https://doi.org/10.1175/1520-0450(1993)032<0884:dopaum>2.0.co;2

Chollet, C. F., et al. (2015). Keras. Retrieved from https://keras.io

Das, S., Datta, S., & Shukla, A. K. (2020). Detection of thunderstorm using Indian navigation satellite NavIC. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(5), 3677–3684. https://doi.org/10.1109/tgrs.2019.2960035

Dolan, B., Kollias, P., van den Heever, S. C., Rasmussen, K. L., Oue, M., Luke, E., et al. (2023). Time resolved reflectivity measurements of convective clouds. *Geophysical Research Letters*, *50*(22), e2023GL105723. https://doi.org/10.1029/2023gl105723

Fiolleau, T., & Roca, R. (2013). An algorithm for the detection and tracking of tropical mesoscale convective systems using infrared images from geostationary satellite. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(7), 4302–4315. https://doi.org/10.1109/TGRS.2012.2227762

Gómez, J. L., Pastoriza, F. T., Fariña, E. A., Oller, P. E., & Álvarez, E. G. (2020). Use of a numerical weather prediction model as a meteorological source for the estimation of heating demand in building thermal simulations. *Sustainable Cities and Society*, *62*, 102403. https://doi.org/10.1016/j.scs.2020.102403

He, J., & Liu, K. (2023). Evolution mechanisms modeling of cloud-ice, cloud-water, and rain from microwave remote sensing. In *IGARSS 2023-2023 IEEE international geoscience and remote sensing symposium* (pp. 1830–1833). IEEE.

Heikenfeld, M., Marinescu, P. J., Christensen, M., Watson-Parris, D., Senf, F., van den Heever, S. C., & Stier, P. (2019). Tobac 1.2: Towards a flexible framework for tracking and analysis of clouds in diverse datasets. *Geoscientific Model Development*, *12*(11), 4551–4570. https://doi.org/10.5194/gmd-12-4551-2019

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, *1*, 278–282.

Huffman, G., Bolvin, D., Braithwaite, D., Hsu, K., Joyce, R., & Xie, P. (2014). *Integrated multi-satellite retrievals for GPM (IMERG), version 4.4*. NASA's Precipitation Processing Center. Retrieved from ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata/

Ito, M., & Masunaga, H. (2022). Process-level assessment of the iris effect over tropical oceans. *Geophysical Research Letters*, *49*(7), e2022GL097997. https://doi.org/10.1029/2022gl097997

Jacobson, R., Shao, X. M., & Holzworth, R. H. (2011). Satellite triangulation of thunderstorms, from fading radio fields synchronously recorded on two orthogonal antennas. *Radio Science*, *46*(6), 1–17. https://doi.org/10.1029/2011rs004783

Joyce, R. J., Janowiak, J. E., Arkin, P. A., & Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, *5*(3), 487–503. https://doi.org/10.1175/1525-7541(2004)005<0487:camtpg>2.0.co;2

Kaur, P., Eriksson, V., Barlakas, S., & Pfreundschuh, S. F., & Fox, S. (2022). Fast radiative transfer approximating ice hydrometeor orientation and its implication on IWP retrievals. *Remote Sensing*, *14*(7), 1594. https://doi.org/10.3390/rs14071594

Kukurugya, J., Terpák, A., & Voigt, A. (2011). Comparison of a simulation model of heat exchanger with real equipment. In *2011 12th International Carpathian Control Conference (ICCC)* (pp. 234–238). IEEE.

Leutwyler, D., Lüthi, D., Ban, N., Fuhrer, O., & Schär, C. (2017). Evaluation of the convection-resolving climate modeling approach on continental scales. *Journal of Geophysical Research: Atmospheres*, *122*(10), 5237–5258. https://doi.org/10.1002/2016jd026013

Li, X., Yang, J., Wang, J., Huang, F., Fang, H., Han, G., & Li, W. (2023). Tropical cyclone winds retrieval algorithm for the Cyclone Global Navigation Satellite System mission. *IEEE Geoscience and Remote Sensing Letters*, *20*, 1–5. https://doi.org/10.1109/lgrs.2023.3318187

Marinescu, P. J., van den Heever, S. C., Grant, L. D., Bukowski, J., & Singh, I. (2024). How much convective environment subgrid spatial variability is missing within atmospheric reanalysis datasets? *Authorea Preprints*, *51*(24), e2024GL111856. https://doi.org/10.1029/2024gl111856

Mastro, P., Masiello, G., Serio, C., Cimini, D., Ricciardelli, E., Di Paola, F., et al. (2022). Combined IASI-NG and MWS observations for the retrieval of cloud liquid and ice water path: A deep learning artificial intelligence approach. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 3313–3322. https://doi.org/10.1109/jstars.2022.3166992

May, J., Hu, L., Ritchie, E. A., Harandi, M., & Tyo, J. S. (2024). Automated segmentation of tropical cyclone clouds in geostationary infrared images. *IEEE Geoscience and Remote Sensing Letters*, *21*, 1–5. https://doi.org/10.1109/lgrs.2024.3358733

Moraes, F. D., Aquino, F. E., Mote, T. L., Durkee, J. D., & Mattingly, K. S. (2020). Atmospheric characteristics favorable for the development of mesoscale convective complexes in southern Brazil. *Climate Research*, *80*(1), 43–58. https://doi.org/10.3354/cr01595

Müller, M. D., & Parlow, E. (2013). Numerical weather prediction as a surrogate for climate observations in practical applications. *Theoretical and Applied Climatology*, *111*(3–4), 577–584. https://doi.org/10.1007/s00704-012-0693-z

Oertel, M., Boettcher, H., Joos, H., Sprenger, M., Konow, H., Hagen, M., & Wernli, H. (2019). Convective activity in an extratropical cyclone and its warm conveyor belt – A case-study combining observations and a convection-permitting model simulation. *Quarterly Journal of the Royal Meteorological Society*, *145*(721), 1406–1426. https://doi.org/10.1002/qj.3500

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pfreundschuh, S., Brown, P. J., Kummerow, C. D., Eriksson, P., & Norrestad, T. (2022). GPROF-NN: A neural-network-based implementation of the Goddard profiling algorithm. *Atmospheric Measurement Techniques*, *15*(17), 5033–5060. https://doi.org/10.5194/amt-15-5033-2022

Pielke, R. A., Cotton, W. R., Walko, R. E. A., Tremback, C. J., Lyons, W. A., Grasso, L. D., et al. (1992). A comprehensive meteorological modeling system—RAMS. *Meteorology and Atmospheric Physics*, *49*(1–4), 69–91. https://doi.org/10.1007/bf01025401

Rakesh, V., Haridas, S., Sivan, C., Manoj, M. G., Shaji, A., Christy, A. A., et al. (2024). Thunderstorm-induced ionospheric perturbations observed over the Indian equatorial sector using VHF radar and GNSS data. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–15. https://doi.org/10.1109/tgrs.2024.3381048

Raut, R., Jackson, M., Picel, M., Collis, S. M., Bergemann, M., & Jakob, C. (2021). An adaptive tracking algorithm for convection in simulated and remote sensing data. *Journal of Applied Meteorology and Climatology*, *60*(4), 513–526. https://doi.org/10.1175/jamc-d-20-0119.1

Saleeby, S. M., & van den Heever, S. C. (2013). Developments in the CSU-RAMS aerosol model: Emissions, nucleation, regeneration, deposition, and radiation. *Journal of Applied Meteorology and Climatology*, *52*(12), 2601–2622. https://doi.org/10.1175/jamc-d-12-0312.1

Seto, R., Koike, T., & Kachi, M. (2023). Feasibility of liquid water path estimation over land using satellite-based Ka-band passive microwave data. *IEEE Transactions on Geoscience and Remote Sensing*.

Singh, I., Bukowski, J., Marinescu, P. J., Grant, L. D., & van den Heever, S. C. (2025). How spatially variable are tropical and subtropical convective environments? *Geophysical Research Letters*, *52*(22), e2025GL116613. https://doi.org/10.1029/2025gl116613

Tapiador, F. J., Kidd, C., Hsu, K. L., & Marzano, F. (2004). Neural networks in satellite rainfall estimation. *Meteorological Applications*, *11*(1), 83–91. https://doi.org/10.1017/s1350482704001173

Tapiador, F. J., Kidd, C., Levizzani, V., & Marzano, F. S. (2004). A neural networks–based fusion technique to estimate half-hourly rainfall estimates at 0.1 resolution from satellite passive microwave and infrared data. *Journal of Applied Meteorology and Climatology*, *43*(4), 576–594. https://doi.org/10.1175/1520-0450(2004)043<0576:annftt>2.0.co;2

Troncoso, B. A., Mateo-Cortes, J. A., Flores, M. J., & Tapiador, F. J. (2025). Sequence-based undersampling: An algorithm for managing imbalanced datasets. *PeerJ Computer Science*, *11*, e3078. https://doi.org/10.7717/peerj-cs.3078

van den Heever, S. C. (2021). NASA selects new mission to study storms, impacts on climate models. *NASA Earth*. Retrieved from https://www.nasa.gov/press-release/nasa-selects-new-mission-to-study-storms-impacts-on-climate-models

Vila, D. A., Machado, L. A. T., Laurent, H., & Velasco, I. (2008). Forecast and tracking the evolution of cloud clusters (ForTraCC) using satellite infrared imagery: Methodology and validation. *Weather and Forecasting*, *23*(2), 233–245. https://doi.org/10.1175/2007waf2006121.1

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang, C., Zheng, G., Li, X., Xu, Q., Liu, B., & Zhang, J. (2021). Tropical cyclone intensity estimation from geostationary satellite imagery using deep convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–16. https://doi.org/10.1109/tgrs.2021.3066299

Wang, Y., Fu, Y., Fang, X., & Zhang, Y. (2014). Estimating ice water path in tropical cyclones with multispectral microwave data from the FY-3B satellite. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(9), 5548–5557. https://doi.org/10.1109/tgrs.2013.2290320

Wang, Y. V., Kim, S. H., Lyu, G., Lee, C. L., Lee, G., Min, K. H., & Kafatos, M. C. (2022). Relative importance of radar variables for nowcasting heavy rainfall: A machine learning approach. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–14. https://doi.org/10.1109/tgrs.2022.3231125

Wang, Y. V., Kim, S. H., Lyu, G., Lee, C. L., Ryu, S., Lee, G., & Kafatos, M. C. (2024). Nowcasting heavy rainfall with convolutional long short-term memory networks: A pixelwise modeling approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 8424–8433. https://doi.org/10.1109/jstars.2024.3383397

Zhang, Y., Wistar, S., Li, J., Steinberg, M. A., & Wang, J. Z. (2016). Severe thunderstorm detection by visual learning using satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(2), 1039–1052. https://doi.org/10.1109/tgrs.2016.2618929

Zheng, G., Liu, J., Yang, J., & Li, X. (2019). Automatically locate tropical cyclone centers using top cloud motion data derived from geostationary satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(12), 10175–10190. https://doi.org/10.1109/tgrs.2019.2931795

Zhuge, X. Y., Guan, J., Yu, F., & Wang, Y. (2015). A new satellite-based indicator for estimation of the western North Pacific tropical cyclone current intensity. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(10), 5661–5676. https://doi.org/10.1109/tgrs.2015.2427035