SII-NowNet: A machine learning tool for nowcasting convection initiation and intensification in the Tropics

Joseph Smith[1], Cathryn E. Birch[1], John Marsham[1,3], David Moffat[2]

[1]School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK

[2]Plymouth Marine Laboratory, Prospect Place, PL1 3DH, Plymouth, UK

[3]UK Met Office, Exeter, EX1 3PB, UK

Corresponding author: Joseph Smith, eejasm@leeds.ac.uk.com

File generated with AMS Word template 2.0

Early Online Release: This preliminary version has been accepted for publication in *Artificial Intelligence for the Earth Systems*, may be fully cited, and has been assigned DOI 10.1175/AIES-D-25-0043.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

ABSTRACT

Nowcasting developing convection is a crucial component of early warning systems in the Tropics. While machine learning has proven effective for radar-based nowcasting, the lack of radar coverage across much of the Tropics creates a significant capability gap. This study presents Simple Initiation and Intensification Nowcasting neural Network (SII-NowNet), a machine learning tool that uses satellite brightness temperatures to produce probabilistic nowcasts of intensifying and initiating convection in the Tropics. SII-NowNet is first demonstrated over Sumatra, Indonesia—a densely populated tropical island with frequent convective activity. For nowcasts of intensifying convection, SII-NowNet outperforms an optical flow model for lead times of 1–6 hours but begins to over-predict events beyond 3 hours, indicating its limit of capability. For nowcasts of initiating convection, SII-NowNet's limit of capability is reached at 2 hours, beyond which it over-predicts events and is outperformed by climatology. SII-NowNet is trained on 8,661 samples (12 months of data), but sensitivity testing shows that the number of samples can be reduced to three weeks for intensification and three months for initiation, before its outperformed by climatology. This has practical implications for the implementation and further development of SII-NowNet in resource-constrained settings. To exemplify generalisability in other Tropical regions, SII-NowNet is tested over New Guinea, Zambia, Congo and West Africa. Without retraining or region-specific tuning, SII-NowNet achieves skill scores comparable to those over Sumatra. Overall, SII-NowNet's promising results, combined with ease of applicability across the Tropics, make it a valuable tool for future operational nowcasting.

SIGNIFICANCE STATEMENT

In the Tropics, short-term forecasts of destructive, rapidly developing storms are critical components of early warning systems—essential for protecting lives and minimizing impacts. However, many regions face challenges in accurate forecasting due to limited access to ground-based meteorological observations. This study presents a machine learning tool that uses freely and continuously available satellite data—covering the entire Tropics—to forecast storm development several hours in advance. The tool performs well in five distinct Tropical regions and requires minimal computational resources, showing its suitability for implementation in resource-constrained settings. Furthermore, the results highlight the challenges in capturing

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

newly initiating storms – a focus for future studies. Continued work on the tool will focus on testing its performance in additional tropical regions.

1. Introduction

Due to its proximity to the Earth's equator, the Tropics receives intense insolation throughout the year. This often leads to the development of deep convection over land, which can initiate and form intense storms on the timescale of hours. For example, in Indonesia, squall lines are typical convective events that initiate over mountainous regions and propagate across the islands over the course of hours, bringing high winds, rainfall, lightning and destruction along their pathway (J. C. Lo and Orton, 2016; Peatman et al., 2023). Numerical weather prediction (NWP) models struggle to capture the rapid onset of convection both spatially and temporally in the Tropics (Ferrett et al., 2021; Porson et al., 2019; Vogel et al., 2020). In 2022, the Early Warnings for All (World Meteorological Organization, 2023) initiative was set up by the World Meteorological Organisation with the aim of protecting the global population from weather-related hazards by 2027. It is therefore necessary to seek alternative and complementary prediction methods to NWP that can provide crucial information for early warning systems of high impact weather.

Nowcasting is the process of using atmospheric observations to generate rapid, short-term predictions (typically up to ~6 hours ahead) of the future atmospheric state (Browning, 1980; Roberts et al., 2022). Unlike NWP, nowcasting is not limited by the requirement of solving large sets of complex numerical equations to forecast the atmosphere. Instead, nowcasting requires computationally efficient predictive techniques to produce quick, short-term forecasts.

Weather radar observations are considered the best way to identify the location and intensity of storms over large parts of a country (Saltikoff et al., 2019), which is the necessary input to nowcasting systems. Traditionally, a technique known as optical flow (Horn and Schunck, 1981) is applied to radar data to produce nowcasts. Optical flow methods identify precipitation features and their trajectories through a series of images, and these are extrapolated along the trajectory using Lagrangian advection to generate future predictions (Pulkkinen et al., 2019). These methods show skill at predicting the propagation of convection; however, they are limited to information in the latest radar image, meaning they cannot predict the development of new convective cells. Nonetheless, many countries in the mid-latitudes, which are largely covered by a weather radar network, use Lagrangian advection for operational nowcasting of convective weather (Bowler et al., 2006; Germann and Zawadzki, 2002; Haiden et al., 2011; Marcos, 2015). Lagrangian advection has also shown value for nowcasting in the Tropics (Hill et al., 2020; Burton et al., 2022; Smith et al., 2024), but produces a drop in skill during the afternoon/evening period when convection begins to initiate and grow.

More recently, Machine Learning (ML) has been the dominant driver of nowcasting developments (McGovern et al., 2023). The most successful studies have trained deep learning models to generate the next radar image based on a time series of the preceding radar images (Shi et al., 2017; Ayzel et al., 2020; Ravuri et al., 2021; Zhang et al., 2023; Wang et al., 2024). Commonly, these studies present a holistic approach in which an entire field of convection is predicted at once to capture all elements of the lifecycle (e.g., propagation, initiation, growth, decay, etc.), demonstrating their performance with case studies of evolving mature storms. Although these approaches continue to gain skill, there is little understanding on the performance of ML-based nowcasting tools in capturing less frequent (and often more challenging) convective processes such as initiation, which is crucial for decision-making (World Meteorological Organization, 2023). To address this key gap in ML nowcasting capability, a logical approach is to isolate these lifecycle elements and focus efforts on individually predicting their complex and chaotic behaviour.

In the Tropics, where convective weather dominates (Schumacher and Rasmussen, 2020), vast regions have unreliable radars or no radar network at all (Permana et al., 2019; Roberts et al., 2022). Geostationary satellite-based observations do, however, provide a continuous stream of data for monitoring current cloud activity from above (Line et al., 2016) – satellite data has also been a key focus in developing rainfall retrieval tools e.g. over Africa (Amell et al., 2025; Hill et al., 2020). For many regions of the Tropics, satellite-based solutions are the only option for nowcasting (Roberts et al., 2022; Youds et al., 2021), and have demonstrated vast improvements in capability (Fletcher et al., 2023; Marcos, 2015). In particular, ML studies in the Tropics have made use of these satellite data observations as inputs to their nowcasting models (Han et al., 2019; Lagerquist et al., 2021; Lee et al., 2017) and proven to be effective approaches. However, they still require radar data for training, resulting in a continued nowcasting gap in the 'radar-less' regions of the Tropics. Nonetheless, Lagerquist et al., (2021) showed particularly good skill when testing U-Net model architectures (Ronneberger et al.,

2015) for nowcasting fields of convection (identified in radar), providing motivation for the further application of these ML architectures.

This study presents Simple Initiation and Intensification Nowcasting neural Network (SII-NowNet), a U-Net model that produces individual nowcasts of convection initiation and convection intensification in the Tropics, with lead times of 1 – 6 hours. The simple training routine and low computational demand of U-Net’s make them a favorable choice for development in resource-constrained settings (e.g. regions in the Global South). Furthermore, SII-NowNet’s sole requirement of brightness temperature data (from geostationary satellites) for training and testing means that it can be easily applied to any region in the Tropics. Overall, SII-NowNet is unique in that it focusses on the prediction of convection initiation and intensification, whilst being a practical tool for implementation and development in the Tropics.

The performance of SII-NowNet is first evaluated over Sumatra, Indonesia – one of the largest Tropical islands in terms of size and population, which experiences regular convective activity (Mori et al., 2004). SII-NowNet is evaluated against established nowcast methods to identify a limit of capability, and the minimum number of training samples required before losing skill is quantified. A feature importance test is also applied to gain insight into SII-NowNet’s decision-making. Finally, the evaluation of SII-NowNet is extended over New Guinea, Zambia, Congo and West Africa to exemplify its generalisability to other regions in the Tropics.

2. Data and Methods

a. Data

To evidence generalisability across the Tropics, SII-NowNet was applied to the five domains shown in Figure 1. SII-NowNet was firstly trained and tested over Sumatra, Indonesia (domain A, Figure 1) and then applied to the remaining four domains for further testing (with no re-training or region-specific tuning).

For the Sumatra and New Guinea (B, Figure 1) domains, SII-NowNet was trained/tested using hourly brightness temperature (BT) data from channel 13 of the Advanced Himawari Imager onboard the Himawari-8/9 geostationary satellite (Bessho et al., 2016) from 01/01/2018 to 31/12/2019 (17,327 hourly images). It detects infrared radiation at a wavelength of 10.4 µm, which allows clear identification of clouds high in the atmosphere by their relatively colder temperatures (Figure 2a). The detected satellite images are transformed to a Cartesian grid with a native grid spacing of 2 km, using the gdalwarp command from the Geospatial Data Abstraction Library (Rouault et al., 2023). As New Guinea is used as an exemplar region, no training is applied over this region and a shorter test period of data is used from 01/03/2019 to 31/03/2019 (740 hourly images).

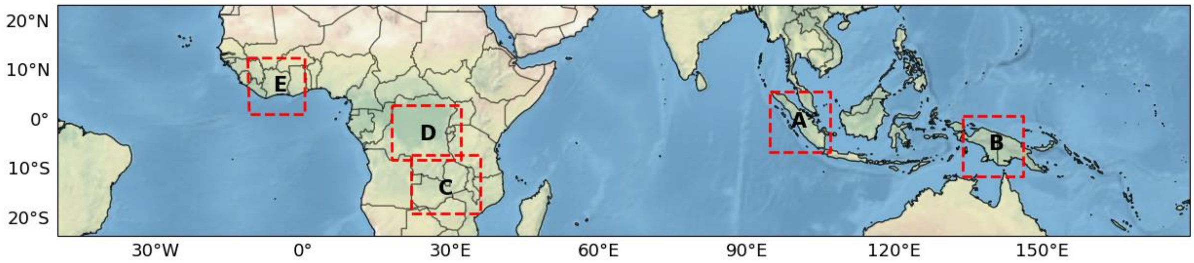

Fig 1. The five domains used in this study. SII-NowNet is first trained and tested on Sumatra, Indonesia (A), then further tested over New Guinea (B), Zambia (C), Congo (D) and West Africa (E).

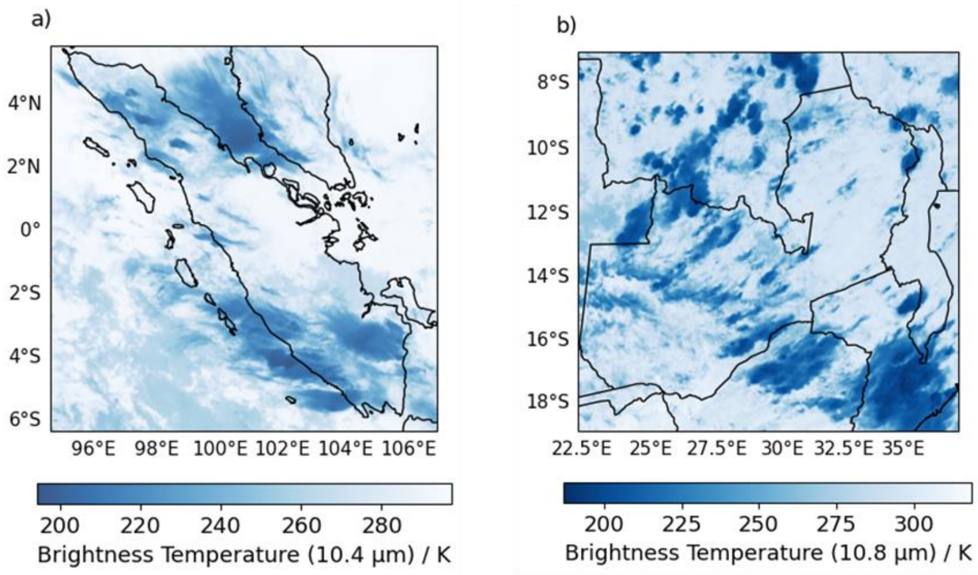

Fig 2. Examples of infrared brightness temperature images over the two domains used in this study. a) Provides an example of a 10.4 µm BT image over Sumatra, Indonesia on 7th March 2019 at 2000 LT. b) Provides an example of a 10.8 µm BT image over Zambia on 15th March 2024 at 1500 LT.

For further testing (without any re-training) over Zambia (domain C), Congo (domain D) and West Africa (domain E), BT data is used from channel 9 of the Spinning Enhanced Visible and Infrared Imager onboard the Meteosat Second Generation (MSG) satellite. The

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

MSG satellite detects infrared radiation at a wavelength of 10.8 µm, every 15 minutes (Schmetz et al., 2002) over each domain (Figure 2b shown for Zambia). Although slightly longer than the 10.4 µm wavelength onboard the Himawari-8/9 satellite, 10.8 µm is the closest available wavelength. The detected 10.8 µm BT images are presented on their native grid spacing of 3 km. As these domains are used as exemplar case studies, a shorter test period is used from 01/03/2024 to 31/03/2024 (2,782 images).

b. Methods

1) SII-NowNet nowcast setup

SII-NowNet requires a 416 x 416 x 2 pixel input matrix (Figure 3). The first two dimensions represent the spatial shape of the input (i.e. 2-D fields covering the target domains), and the third dimension represents the two predictor channels. The two predictor channels are:

1. BT T-0 hrs
2. BT T-1 hrs

where T represents the nowcast initialisation time. BT T-0 provides information on present atmospheric conditions, whereas BT T-1 provides information on conditions one hour before present. By learning the difference between BT T-0 and BT T-1, SII-NowNet extracts information on how conditions have changed over the previous hour leading up to the initialisation time i.e. identifying whether convection is starting to develop or decay (note, the addition of BT T-2 hours was tested but provided no performance gain). Using only two BT input fields, spaced 1 hour apart, provides the balance between reducing computation and training time, whilst also allowing changes in cloud conditions to be captured.

SII-NowNet uses the input matrix to produce a 26 x 26 gridded nowcast over the same domain (see next section for dimension justification). i.e., it spatially coarsens the input data by a factor of 16 (Figure 3). The grid size of the nowcast therefore depends on the gridded resolution of the geostationary satellite data. For example, using Himawari-8/9 BT data with a grid box spacing of 2 km results in a SII-NowNet nowcast with a 32 km grid box resolution, covering a region of 832 km x 832 km (94.6°E – 106.9°E, 6.4°S – 5.9°N).

Each grid box in a SII-NowNet nowcast is assigned a value between 0 and 1, representing the probability of a convective event occurring over that grid box for a given lead time. This study isolates two types of convective events for nowcasting:

1. Nowcasting the probability of convection intensification at each lead time
2. Nowcasting the probability of convection initiation at each lead time

SII-NowNet is trained separately for intensification and initiation, producing nowcasts with lead times at hourly increments for both. Predictions are made on hourly timescales to i) test SII-NowNet on the largest temporal scales in nowcasting and ii) match the temporal scale of the BT inputs. For each initialisation time, SII-NowNet produces a total of 12 nowcasts, six for intensification and six for initiation (representing lead times of 1 to 6 hours).

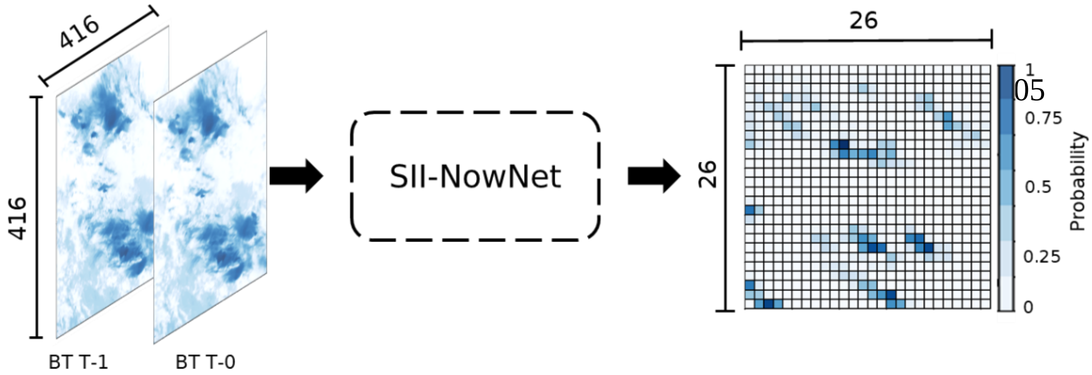

Fig 3. A diagram of the SII-NowNet nowcast process (applicable to both intensification and initiation predictions). The 416 x 416 x 2 input matrix (left) is input to SII-NowNet (middle), which outputs a 26 x 26 probabilistic nowcast over the same domain as the inputs (right).

2) Convective event observations

To generate specific nowcasts of initiation and intensification, SII-NowNet must first be trained on a dataset of observed events for each type (initiation and intensification). One of the aims of this study is to use solely satellite data, as a solution to the unreliability/unavailability issue of radar data in the Tropics. Therefore, a method for identifying intensification and initiation events in historical satellite observations is presented.

To identify intensification and initiation events, this study uses a set of empirically tested thresholds applied to BT imagery. The use of pre-defined thresholds for identifying convective events is a common approach. The simplest approaches use a single ≥35dBZ threshold criteria, applied to radar reflectivity pixels, to identify them as initiating or non-initiating (Han et al., 2017; Lee et al., 2017). More advanced methods apply several criteria to first identify and track existing convective regions before separating them from the new initiations (Han et al., 2019; Mecikalski et al., 2015).

In this study, both intensification and initiation events are identified on a 26 x 26 gridded domain (hence the size of the SII-NowNet nowcast). This grid sizing was chosen through empirical testing, as it was found to be the most appropriate scale for accurately capturing initiation and intensification events.

(i) Intensification identification

Figure 4 provides an example of intensifying convection events over Sumatra that are identified using the method. For each individual grid box (within the 26 x 26 gridded domain), convection intensification is identified by comparing two consecutive BT images, separated by an hour (T-1 hours and T-0 hours, where T is the time of the intensification event), over the grid box. Each BT pixel within the grid box is defined as intensifying if it satisfies two criteria:

1. BT decreases by >= 20 K between T-1 and T-0
2. BT <= 235 K at T-0

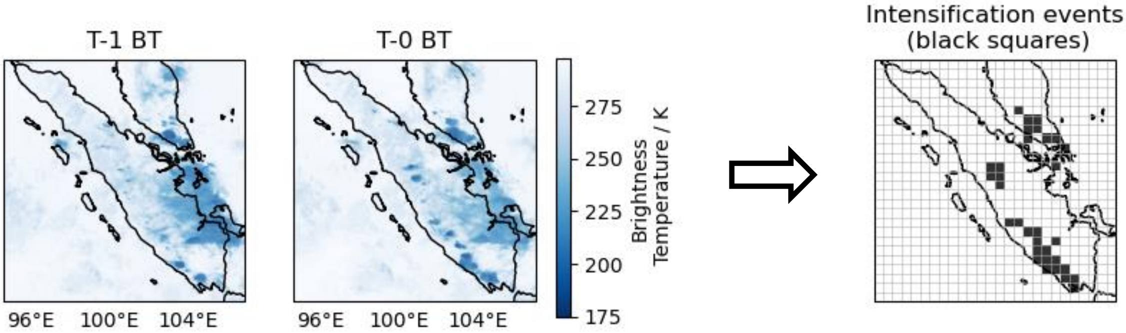

Fig 4. An example of intensifying convection along the south coast of Sumatra on 5th January 2019 at 1800LT. By comparing the T-1 BT and T-0 BT (left of arrow), the intensification identification method extracts only the intensifying convection, which it outputs on a 26 x 26 gridded domain (right of arrow).

A threshold of 235 K is used in step 2 to try and isolate as much convection as possible, and is a common choice for the tropics in the literature (Crook et al., 2024; Feng et al., 2021; Machado and Laurent, 2004; Roca et al., 2017). Finally, in order to account for BT pixel noise, if the number of intensifying BT pixels within the grid box is >=5%, the grid box is labelled 1 for intensifying or 0 for not intensifying. This process is simultaneously repeated for each grid box in the 26 x 26 gridded domain and then applied to all available BT image timesteps. This

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

results in 17,326 intensification gridded samples for Sumatra, 740 for New Guinea and 2,782 for Zambia, Congo and West Africa.

The definition of intensification described in this study is local to each grid box i.e. it is within a Eulerian framework, meaning that it is not “aware” of convective systems and their dynamics. Therefore, by this definition, an intense storm moving over a previously cloud-free grid box is classed as intensification (as well as stationary intensification over a grid box).

(ii) Initiation identification

Convection initiation is also identified using two consecutive BT images, separated by an hour (T-1 hours and T-0 hours, where T is the time of the initiation event). To be classified as an initiation, the new convective cells have to be isolated from the convection that is already present (at T-1). The method follows two key steps, which are applied to each BT pixel within the nowcast domain (Figure 5 provides an accompanying example of the method):

1. The growth/propagation of convection present within the T-1 BT is simulated to produce an

estimate of its size in the T-0 BT

2. The estimated T-0 BT is subtracted from the observed T-0 BT to leave the remaining

new convective cells that develop over the hour i.e. the convection initiations.

The convection growth simulation is performed by thresholding the T-1 BT at <=255 K and then applying a smoothing filter (Figure 5a), which ‘grows’ the convection to account for its potential expansion and/or propagation at speeds ≤11 m/s. This maximum speed is in agreement with peak storm speeds measured by Mori et al., (2004). A 255 K threshold includes the regions of cold cloud <=235 K (to remain consistent with the intensification identification), plus an additional 20 K allowance for regions where the convection may deepen and decrease in temperature by 20 K over the following hour. This additional 20 K allowance is known as the growth allowance. The T-0 BT observation is thresholded at 235 K to isolate only the convective regions (Figure 5b). The T-0 BT observation is then subtracted from the T-0 BT estimated field (Figure 5c).

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

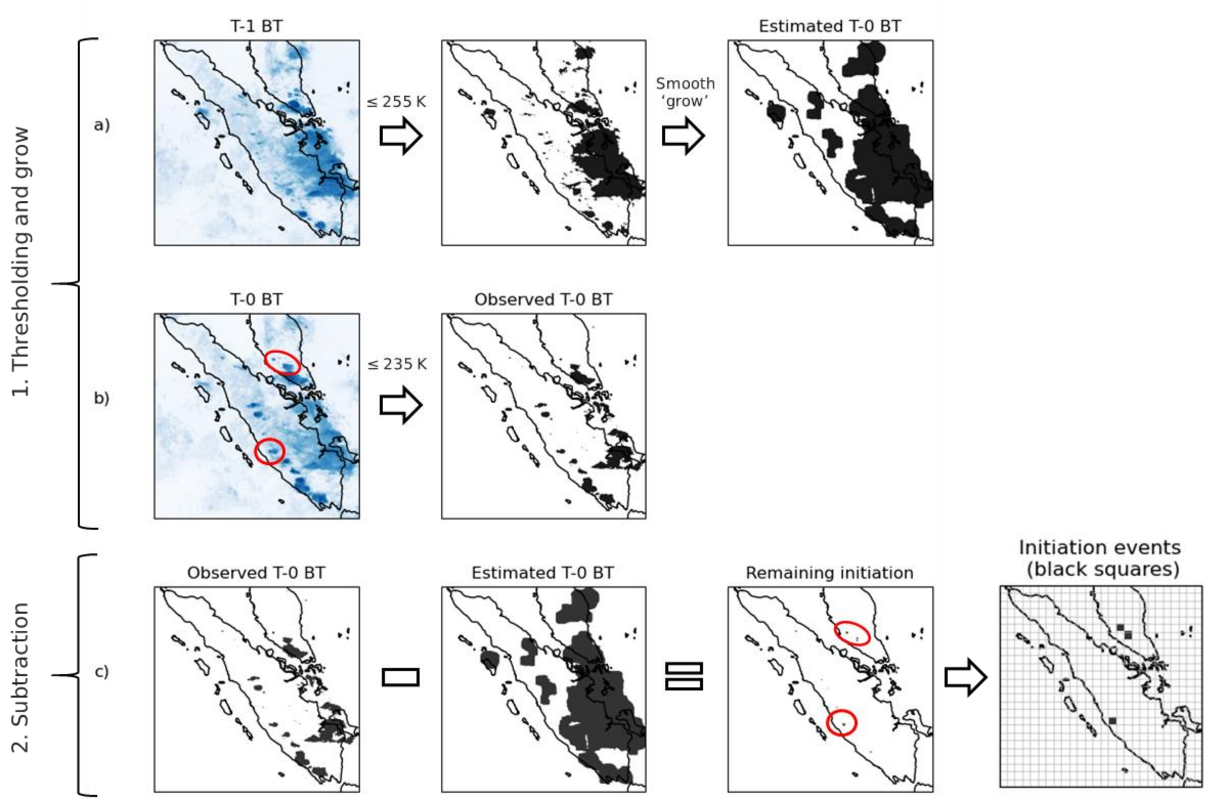

Fig 5. Using the same case as Figure 4, the two key processes for identifying convection initiation. a) The thresholding and then smoothing of the T-1 field to 'grow' the convection that is already present (T-0 BT estimate). b) Thresholding of the T-0 field to extract the convective regions only (T-0 BT observation). c) The final step showing the T-0 BT estimate field being subtracted from the T-0 observation field to leave only the isolated initiation (circled in red).

This method assumes that the present convection in the T-1 BT will propagate or grow between T-1 and T-0. However, if in reality the convection does not evolve in this way (e.g. it dissipates), then subtracting the T-0 BT estimate from the T-0 BT observation will result in regions with negative values. To overcome this, all negative values are converted to 0, representing no convection initiation.

Finally, in order to account for BT pixel noise, for each grid box in the 26 x 26 gridded domain, if the number of initiation BT pixels is >=5%, the grid box is labelled as 1 for initiating or 0 for not initiating. This process is simultaneously repeated for each grid box in the 26 x 26 gridded domain and then applied to all available BT image timesteps. This results in 17,326 intensification gridded samples for Sumatra, 740 for New Guinea and 2,782 for Zambia, Congo and West Africa.

The initiation identification method relies on applying a set of BT thresholds to isolate the newly developing convection. Figure 6 shows how the number of identified initiations decreases when the growth allowance value is varied from 0 K – 30 K. As the growth allowance is increased, more isolated initiations are encompassed by the T – 0 BT estimate (Figure 5a), reducing the overall number of initiations. The rapid decrease shows that the method is highly sensitive to the growth allowance value, a limitation acknowledged by the authors.

The key aim of this proof-of-concept study is to show that ML can be used to predict initiation/intensification events with skill. In order to do this, a set of pre-defined thresholds must be used to generate datasets of these observed events. Through empirical testing, thresholds were chosen that produced the most useful outcomes for identifying initiation/intensification events. Further development of the initiation/intensification identification methodologies is therefore considered beyond the scope of this study.

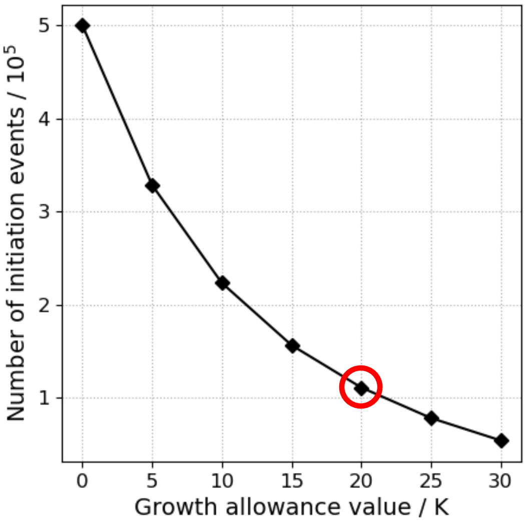

Fig 6. The number of events that are identified by the convection initiation identification methodology as the growth allowance value is varied. The red circle indicates the number of initiations for the growth allowance value that is used in this study.

(iii) Climatology of identified convective events

Table 1 shows the total number of observed events and non-events (for each grid box) that are identified for initiation and intensification over Sumatra. Initiation events (1% of dataset)

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

are observed less frequently than intensification events (14% of dataset). Both intensification and initiation have a clear diurnal cycle in their frequency, where event activity is minimal between 0000 LT – 1000 LT (Figure 7). Initiation events show a sharp peak in frequency at 1400 LT that quickly declines into the afternoon, whereas intensification events are at maximum frequency approximately 3 hours later and continue to maintain significant signal during the afternoon and evening. Peak activity in the afternoon for both types of event is typical in many regions of the tropics (Mori et al., 2004; Yang and Slingo, 2001).

Dataset	Events	Non-events	Balance
Intensification	1,652,207	10,060,169	14%
Initiation	110,838	11,601,538	1%

Table 1. The total count of observed events/non-events for both convection intensification and initiation over Sumatra, Indonesia using hourly brightness temperature over the study period (01/01/2018 – 31/12/2019).

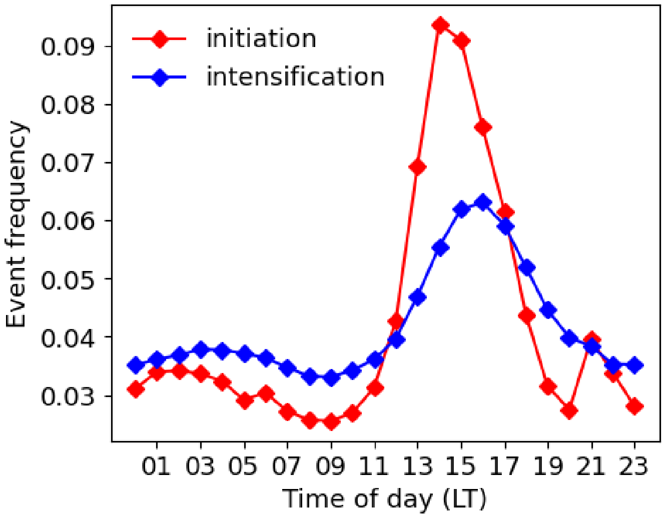

Fig 7. The frequency of observed events for convection intensification and convection initiation (Table 1) at each hour of the day over Sumatra.

3) SII-NowNet architecture

The SII-NowNet model is a deep learning model based on a U-Net architecture (Ronneberger et al., 2015), named due to the U-shaped path that the input takes through the

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

model (Figure 8). The 416 x 416 x 2 input matrix undergoes four encoding and decoding stages before processing into the desired output.

During the first part of each encoding stage, the input (represented by the black arrows in Figure 8) is passed through two 2-D convolutional layers, each with a 3 x 3 filter, which aim to identify key features in the data. Attached to the end of each 2-D convolutional layer is a batch normalization function (to stabilize the network during training) and a ReLU activation function (to enable non-linear patterns in the data to be learnt). In the second part of the encoder stage, the output from the two convolutions is passed through a max pooling function with a 2 x 2 filter. This downscales the output whilst retaining the most important features, resulting in a more compressed representation of the input data. However, downscaling can cause important information to be lost. To account for this, the output from the first part of the encoding stage (before max pooling) is saved and transferred back into the model at a later stage (during decoding). This is known as the skip connection and helps to retain fine spatial details. After 4 encoder stages the spatial dimensions of the original input data have been compressed to a 26 x 26 grid, and the dimension of the third channel is increased to 1024. This 26 x 26 x 1024 matrix is known as the latent space and represents the most important features that the model has identified from the input.

To represent the latent space in a simple way (i.e. on a 2-D grid), it must be decoded. In each decoding stage, a 2-D transposed convolution (with a 2 x 2 matrix) is applied to the input matrix, which up-scales the spatial dimensions. Next, the output saved in the corresponding skip connection (from the earlier encoding step) is concatenated to the up-scaled matrix to re-introduce any lost spatial information from encoding. Finally, the resulting matrix is passed through two 2-D convolutional layers (which also have batch normalization and ReLU activation functions attached).

After 4 decoding stages the matrix has the dimension 416 x 416 x 64. To reach the desired dimension of 26 x 26 x 1, the matrix is passed through a 2-D convolutional layer (with a 1 x 1 filter matrix) and then a sigmoid activation function. The sigmoid activation function ensures the output is between 0 and 1, representing the model’s probabilistic prediction. Finally, a max pooling layer (with a 16 x 16 matrix) is used to transform the output into a 26 x 26 x 1.

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

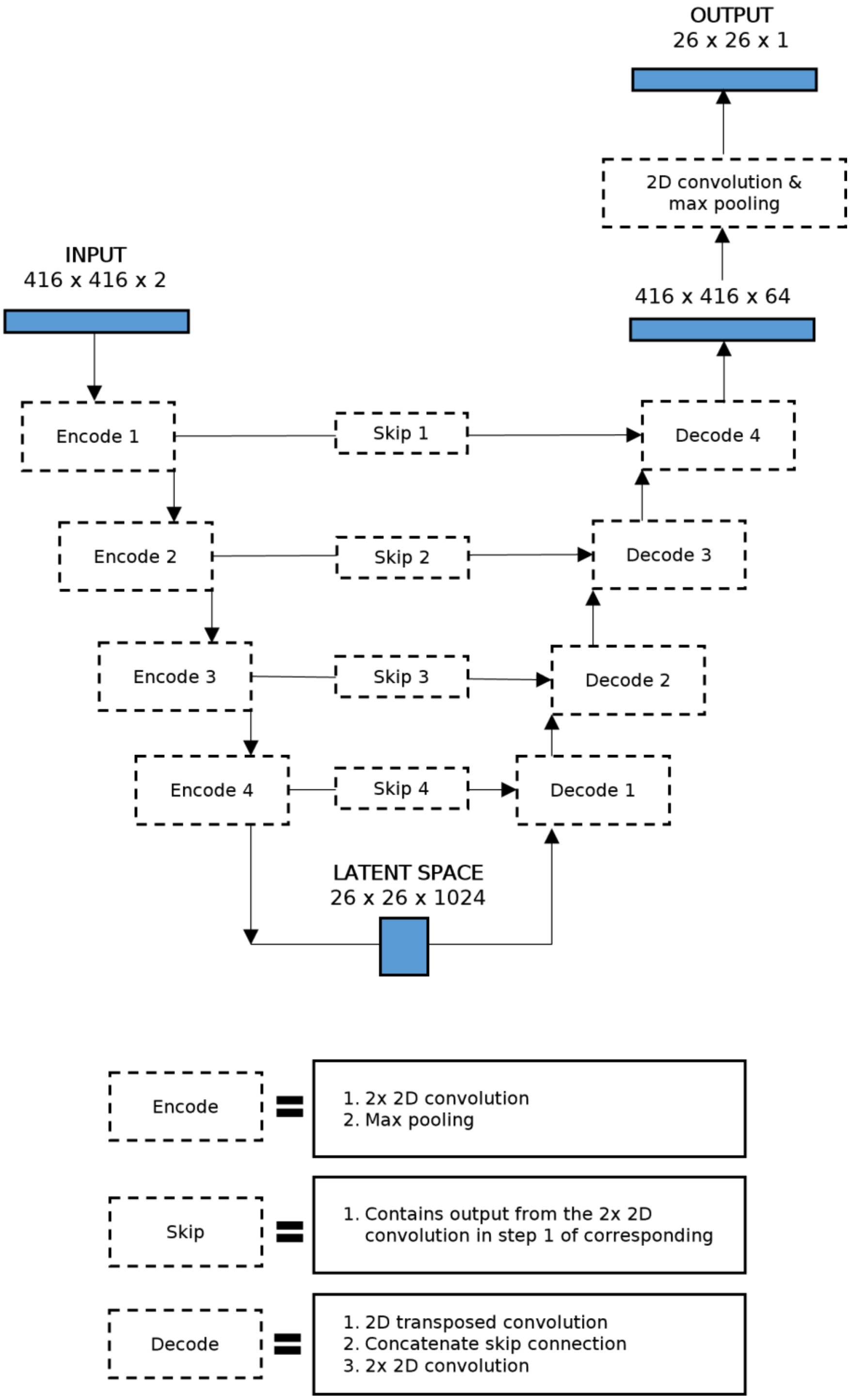

Fig 8. A diagram of the SII-NowNet model. A U-Net architecture with four encoder steps followed by four decoder steps, with skip connections in between. A final processing stage converts it to the correct dimension. The arrows represent the input matrix entering each step. Each blue rectangle shows the shape of the input matrix as it passes through different stages of the network.

4) Training and testing data

For each hourly lead time, for both intensification and initiation nowcasts, a unique version of SII-NowNet was trained (resulting in 12 unique SII-NowNet models presented in this paper). Each model was trained for an upper limit of 10 epochs (i.e. 10 full passes through the entire training dataset) using a binary cross entropy loss function with the Adam optimizer (learning rate = 0.001). However, the maximum number of required training epochs across all lead times for intensification and initiation was 8. To ensure no overfitting to the training data, an early stopping routine was applied to the validation loss function (a 20% random subset of the training data was used as the validation data). If the loss value stayed above the lowest previous loss value for more than 1 epoch, training ended and the model returned to its state at the previous lowest loss value.

For training SII-NowNet over Sumatra, the available 17,326 initiation and intensification samples were divided into a training set and a test set. 8,661 training samples were taken from 2018 and used during the training process for model learning, whilst 8,665 test samples were taken from 2019 and were used for model verification only. Therefore, the test samples were unseen by the model and excluded from training.

5) Verification

(i) Baseline models

In this study, we compare SII-NowNet to three baseline models: Short-Term Ensemble Prediction System (STEPS), climatology and persistence.

STEPS is a probabilistic nowcasting tool that can be used to produce predictions on 1 – 6 hours’ lead time (Bowler et al., 2006). It is an established nowcasting method that is currently used in many operational settings (e.g. UK Met Office), making it an appropriate baseline comparison. STEPS employs optical flow (Horn and Schunck, 1981) to estimate the motion of existing convection and then extrapolates along the motions vectors using Lagrangian advection. Furthermore, STEPS aims to account for the uncertainty in the convection by injecting various intensities of noise at different length scales (higher noise intensity at smaller, more unpredictable scales) and temporal scales (higher noise intensity at longer, more unpredictable lead times). By using new realizations of the noise fields, STEPS is able to produce an ensemble of nowcasts. A more detailed description of STEPS is provided by

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

(Bowler et al., 2006). This study uses the open-source pySTEPS library (Pulkkinen et al., 2019) implementation of STEPS to produce 20-member ensemble nowcasts of BT for 1 – 6 hours' lead time, following the methods in Smith et al., (2024). The initiation and intensification identification methods (section 2.b.2) are then applied to the STEPS BT nowcasts (i.e. simply replacing the BT observations with the STEPS nowcasts at the same valid time), resulting in corresponding STEPS nowcasts for initiation and intensification.

In this study, climatology is defined as the expected average number of convective events for a given time of day. Climatology is calculated separately for intensification and initiation at each hour of the day by computing the fraction of observed events from the sum of observed events and non-events. Given the large amplitude of the diurnal cycle of convection in the Tropics (Yang and Slingo, 2001), the climatology at each hour is verified only against the set of observations that are valid at the same hour.

A persistence nowcast is a simple prediction method, which assumes that the current weather will persist into the future i.e. the weather now will be identical at the next prediction timestep. This baseline model is produced by using the latest event observation as the next prediction for a given lead time.

(ii) Basic verification statistics

To evaluate SII-NowNet against observations, the probabilistic output produced by the model needs to be thresholded at a chosen probability (a value between 0 – 1) to generate a binary prediction field. Each grid box value in the binary field can then be compared to the corresponding grid box value in the observation. A hit (a) is counted when both grid box values equal 1, a miss (b) is counted when the observation grid box equals 1 and the binary prediction equals 0, a false alarm (c) is counted when the observation grid box value equals 0 and the binary prediction grid box value equals 1, and a correct negative (d) is counted when both grid box values equal 0.

These metrics are used to calculate the accuracy, probability of detection (POD), false alarm ratio (FAR) and the probability of false detection (POFD):

$$accuracy = \frac{a+d}{a+b+c+d} \tag{1}$$

$$POD = \frac{a}{a+b} \tag{2}$$

$$FAR = \frac{c}{a+c} \tag{3}$$

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

$$POFD = \frac{c}{c+d} \tag{4}$$

(iii) ROC curves

Receiver operating characteristic (ROC) curves are an evaluation method used to analyse the skill of probabilistic predictions. Each nowcast is thresholded at 11 equally spaced values (to provide a high enough density of points) within a given probability range. This probability range will vary depending on the range of probabilities produced by the nowcast. For example, a nowcast that produces probabilities between 0 and 1 will be sampled in the range [0,1] in 0.1 increments. The binary outputs are then compared to the corresponding binary observation and the POD and POFD scores are calculated (resulting in 11 scores – one for each threshold). This process is applied to the entire test set and the mean POD is plotted against the mean POFD at each threshold. Finally, this is repeated for all 6 lead times, which produces 6 POD – POFD lines spanning the probability range. A skilful nowcast will maximise the POD and minimise the POFD, which results in a high Area Under the Curve (AUC) score – the key metric.

(iiii) Reliability diagram

Reliability diagrams give insight into how a probabilistic prediction of an event compares to the observed frequency of that event. The nowcast probabilities are binned to create 10 sub-group bins and, for each sub-group bin, the frequency of observed events is calculated. The mean probability within each sub-group bin is then plotted against the corresponding frequency of observed events. For a perfectly reliable nowcast the predicted probability of events will exactly match the observed frequency of events, resulting in a line along the x=y line.

(iiiii) Permutation importance

The permutation importance method provides insight into the input predictors that have the greatest influence on the performance of SII-NowNet. In turn each predictor (for each test case) is replaced by Gaussian noise and input to SII-NowNet. The new AUC score is calculated and then subtracted from the original SII-NowNet AUC score (without noise replacement). This is repeated 20 times for each predictor to get a distribution of AUC score differences. A greater AUC score difference implies that the predictor has a greater influence on SII-NowNet's decision and is therefore an important variable for prediction. A small AUC score difference implies that the predictor has little influence on the SII-NowNet's decision and is therefore a less important variable for prediction.

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

3. Results

a) Qualitative analysis of SII-NowNet

Figure 9 presents a case study for SII-NowNet intensification and initiation nowcasts, valid at 1300 LT on 27th December 2019 over Sumatra, Indonesia. Figure 9a and b show the BT images that were used to identify the initiation/intensification events. This case was chosen because it highlights both the strengths and weaknesses of SII-NowNet. Contouring has been applied to all SII-NowNet nowcasts to provide a smoother probabilistic field (compared to the raw gridded output in Figure 3). Different colour bar scales have been used for intensification and initiation nowcasts to account for their differing probability ranges at each lead time.

At 1 hour lead time, SII-NowNet's intensification (Figure 9d) and initiation (Figure 9k) nowcasts produce regions of high probability that align with the observed intensification (Figure 9c) and initiation (Figure 9j) events e.g. southeast coast of Sumatra for both nowcasts. By producing these distinct regions of high probability, it highlights SII-NowNet's ability (at this lead time) to clearly discriminate between events and non-events. However, the initiation nowcast also provides an example of missing an initiation event (northwest Sumatra coastline, Figure 9j) and a false alarm (northeast Sumatra coastline, Figure 9k), highlighting the challenge in capturing isolated events.

At 2 hours' lead time, SII-NowNet's intensification and initiation nowcasts continue to capture the observed events, but with a drop in confidence. The intensification nowcast (Figure 9e) still produces high probabilities that align with the bulk of observed events (e.g. southeast of Sumatra, Figure 9c), however, it now produces a maximum probability of 70% west of the coast of Sumatra (previously 90%–100% at 1 hour), where events are observed. Similarly, the initiation nowcast (Figure 9l) continues to produce the highest probabilities over north and south Sumatra (similar to Figure 9k), which align with the observations, however, the maximum probability has dropped to 20%. Furthermore, the initiation nowcast contains a larger region of initiation probabilities to the east of the domain, which does not align with the observations (i.e. it produces a false alarm here).

Beyond 2 hours' lead time, SII-NowNet's nowcasts increase in blurriness and decrease in maximum probability. In particular, the initiation nowcasts (Figure 9m-p) show a rapid drop in probability over this period, along with a reduction in precision. The intensification nowcast at 3 hours' lead time (Figure 9f), although blurry, still appears to show some precision, aligning

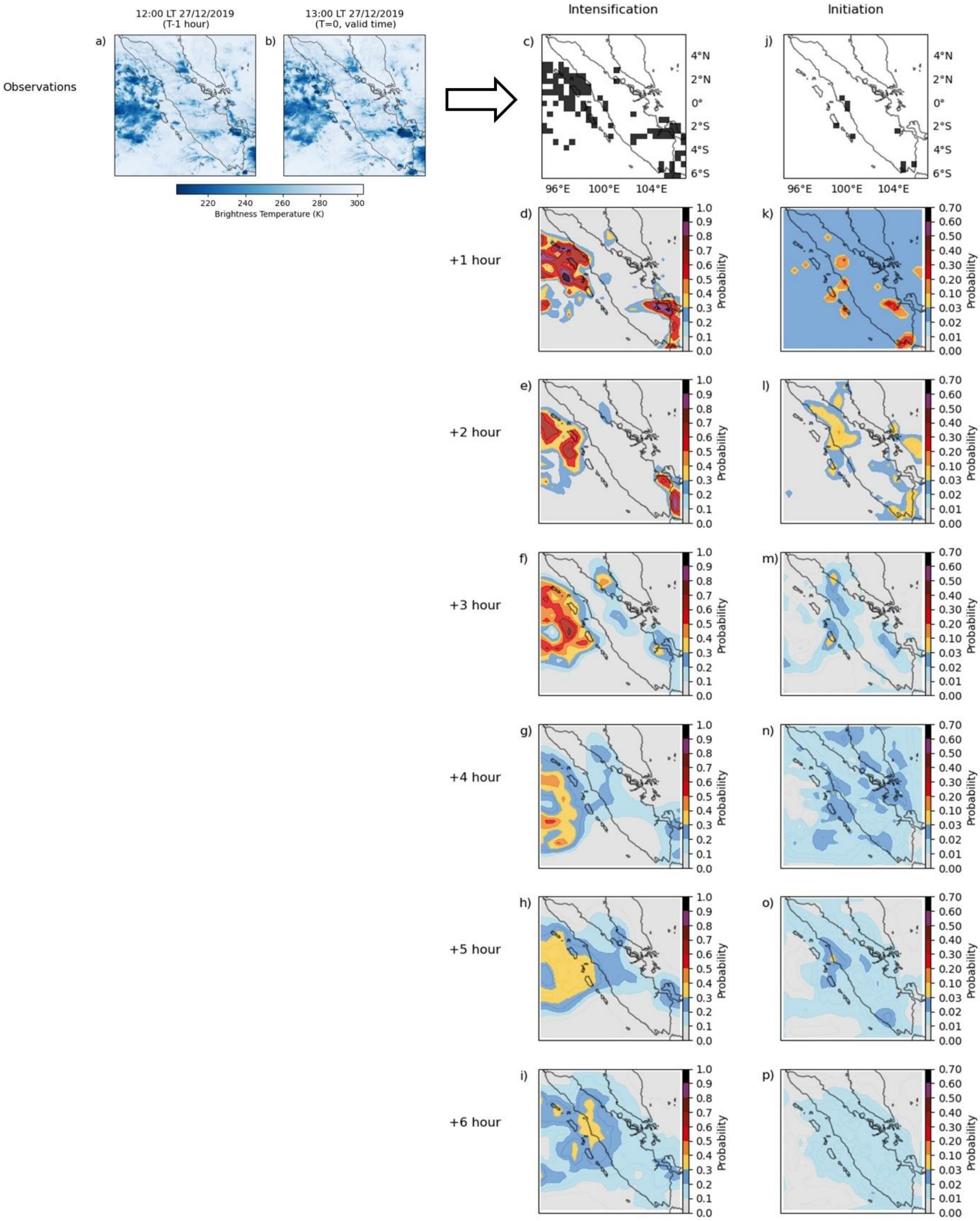

Fig 9. A case study of intensification and initiation events over Sumatra valid at 1300 LT on 27th December 2019. a) and b) show the BT observations at 1200 LT and 1300 LT that were used to identify the events. d-i) and k-p) show the probabilistic SII-NowNet nowcasts for intensification and initiation, respectively, for 1 – 6 hours' lead time. Each nowcast is valid at 1300 LT. Different colour bar scales have been used for intensification and initiation nowcasts to account for their differing probability ranges.

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

with the observed intensification west and southeast of Sumatra. However, for 4 – 6 hours' lead time (Figure 9g-i), the regions of higher intensification probability become much smoother, making it difficult to gauge specifically where convection will intensify.

For initiation nowcasts between 1 – 3 hours' lead time (Figures 9k-m), the minimum background probability reduces from less-than-3% to less-than-1% (before staying consistent beyond 3 hours). SII-NowNet is becoming more uncertain about its predictions of the (far less frequent) initiation events as lead time increases. To compensate for this uncertainty SII-NowNet reduces its probability values (including the background uncertainty).

b) Quantitative analysis of SII-NowNet

The performance of SII-NowNet was quantitatively tested and compared to baseline models on 8,665 samples (equivalent of 12 months) of unseen data over Sumatra from 2019.

Figure 10 shows the ROC curves (a and b) and reliability diagrams (c and d) for SII-NowNet, STEPS and climatology nowcasts for 1, 2 and 3 hour lead times only (to avoid too many lines). The 4 hour lead time intensification result has also been added to the reliability diagram (d) to show its limit of capability. To account for SII-NowNet producing initiation nowcasts with varying probability ranges at different lead times (e.g. Figure 9i-n), their ROC curves are calculated across different probability ranges. Probability ranges of [0, 1], [0, 0.5], [0, 0.1], [0, 0.1], [0, 0.1], [0, 0.1] were used for 1 – 6 hour lead time initiation ROC curves. For intensification nowcasts, all ROC curves were calculated using a probability range of [0, 1].

For intensification, SII-NowNet's ROC curves shift closer to the dashed x=y line (no skill) as lead time increases, resulting in decreasing AUC scores (inset panel Figure 10a) with lead time (0.94 to 0.69). At each lead time shown, SII-NowNet's ROC curves produce a greater AUC score than STEPS, reflecting a better nowcasting performance. At 6 hours' lead time, SII-NowNet's AUC score is only 4% greater than climatology. In terms of reliability, SII-NowNet produces more reliable intensification nowcasts (i.e. closer to x=y) than STEPS for 1, 2 and 3 hours' lead times. At 4 hours' lead time SII-NowNet drastically over-predicts events at probabilities greater than 0.75, generating false alarms. Therefore, although the AUC scores show that SII-NowNet's intensification nowcasts outperform both STEPS and climatology over 1 – 6 hours' lead time, the poorer reliability beyond 3 hours' lead time indicates this as a reasonable limit of capability.

For initiation, the AUC scores (inset panel Figure 10b) produced by the ROC curves are less than the intensification scores for each corresponding lead time i.e. SII-NowNet is less skilful at predicting initiation events. The initiation scores decrease from 0.82 to 0.64 between 1 – 4 hours' lead time. For 4 – 6 hours, SII-NowNet's AUC scores remain approximately constant. SII-NowNet's initiation nowcasts outperform STEPS at each lead time but drops below the climatology between 2 – 3 hours' lead time. This represents a 2 hour lead time limit of capability for SII-NowNet's initiation nowcasts. This limit is also reflected in the reliability curves (Figure 10d), which show that SII-NowNet produces more reliable initiation nowcasts for 1 – 2 hours lead time than STEPS. For 1 hour lead time nowcasts these reliable predictions lie between a probability window of 0 – 0.6. For 2 hour lead time nowcasts the reliability probability window is approximately 0 – 0.15. For predicted probabilities between 0.15 – 0.22, SII-NowNet starts to under-predict initiation events at this lead time (the reliability curve tends towards the y-axis). At 3 hours' lead time, SII-NowNet's initiation nowcasts produce reliability curves that only extend along the x-axis, meaning that it is predominantly producing false alarms.

In general, SII-NowNet produces lower probabilistic values when nowcasting initiation compared to intensification (as exemplified in Figure 9). This suggests that the model is compensating for the relatively lower frequency of initiation events by associating its predictions with more uncertainty i.e. lower likelihood.

SII-NowNet's intensification nowcasts can also be compared against a persistence nowcast, which, on nowcasting timescales, can often have reasonable skill (Smith et al., 2024). Due to the nature of initiations, it is not possible to produce an initiation persistence nowcast (initiations are one-off events that cannot persist), so only a persistence of intensification is presented. At each lead time the SII-NowNet intensification nowcasts have been thresholded using an optimum probability threshold value (0.2, 0.1, 0.2, 0.2, 0.2 and 0.2 for 1 – 6 hours' lead time), to produce a set of deterministic nowcasts. The optimum threshold for each lead time is determined by using the threshold value from the corresponding ROC curve that produces the maximum POD – POFD difference. Figure 11 shows the accuracy, POD and 1–FAR of the deterministic SII-NowNet and persistence nowcasts of convection intensification (1–FAR has been used so that all skill scores of 1 represent a perfect score). In terms of accuracy (Figure 11a), persistence and SII-NowNet perform similarly up to 4 hours, beyond which persistence has a greater accuracy score by approximately 0.1. The persistence nowcast

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

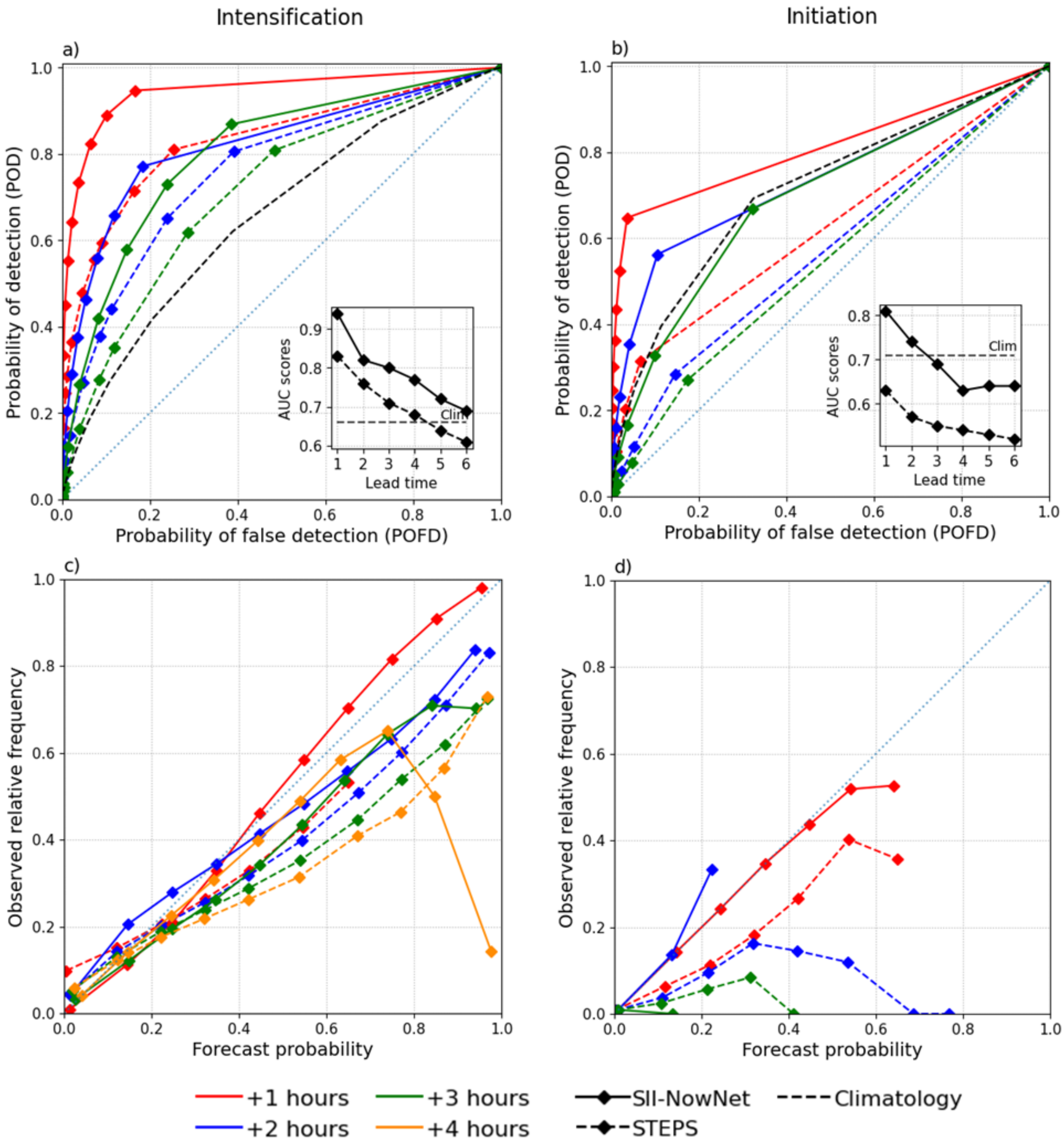

Fig 10. ROC curves for SII-NowNet, STEPS and climatology a) intensification and b) initiation nowcasts, containing curves for each lead time. The inset plots contain the AUC scores at each lead time, with the climatology shown by the horizontal dashed line. Reliability diagrams for SII-NowNet and STEPS c) intensification and d) initiation nowcasts, containing curves for each lead time. By definition the climatology reliability line lies on the x=y line. All analyses are generated using 8,665 unseen samples from 01/01/2019 – 31/12/2019. All plots contain a dashed x=y line for reference.

gains the majority of its accuracy from correct negatives, whereas SII-NowNet's accuracy is derived from a combination of correct negatives and hits. This is shown in the POD scores (Figure 11b), where SII-NowNet outperforms persistence across all lead times by 0.45. Although persistence will often be correct in predicting no event, when an event does happen it is 45% less likely to produce a hit. On the other hand, SII-NowNet has learned the conditions within the predictor fields that are associated with an intensification event, meaning it will more often produce a hit. Even though SII-NowNet produces a greater number of intensification events, it still maintains a greater 1 – FAR score than persistence by approximately 0.05 across all lead times. This means that, when both nowcasts predict an intensification event, persistence is 5% more likely to produce a false alarm than SII-NowNet.

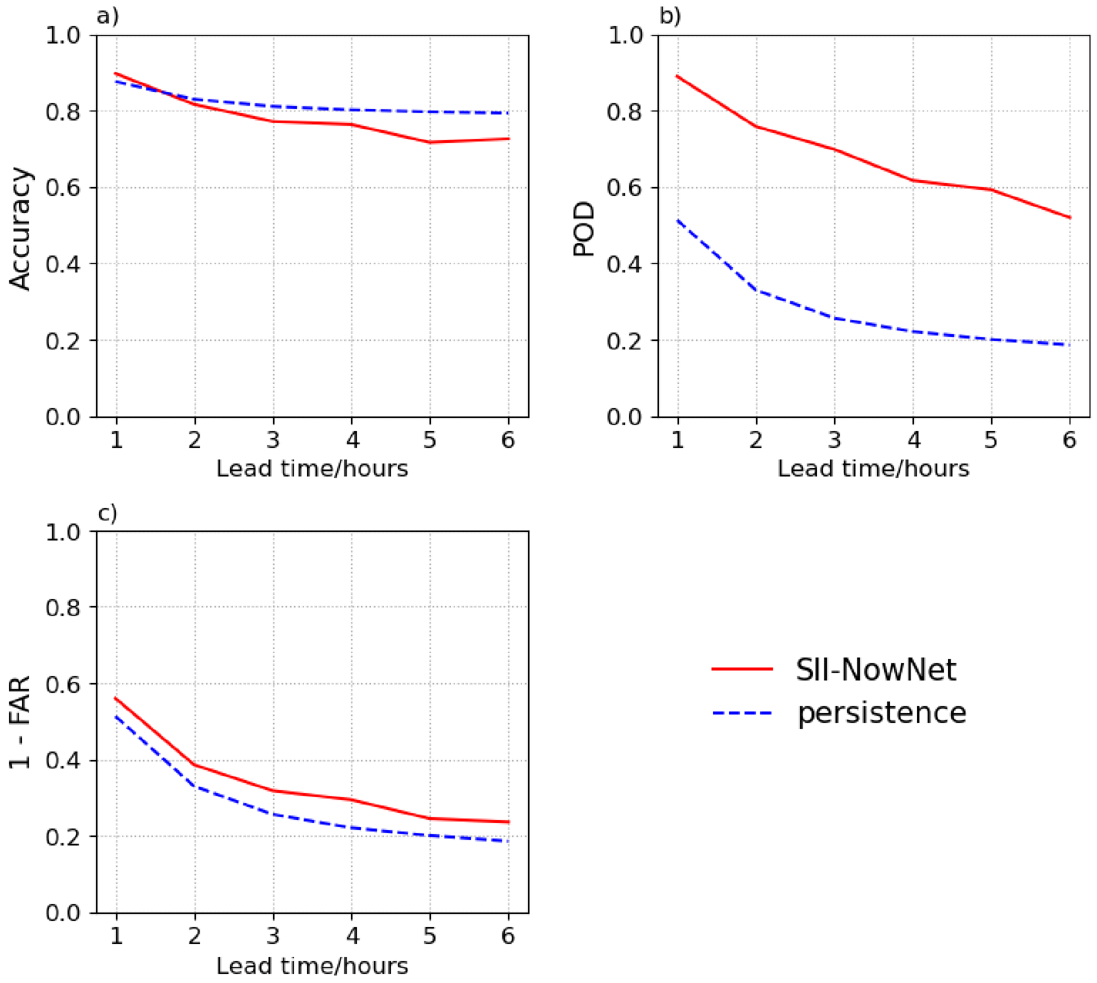

Fig 11. Mean scores for a) accuracy, b) POD and c) 1-FAR across each lead time for intensification nowcasts produced by SII-NowNet and persistence using the same test data as in Fig 10.

c) Sensitivity testing the number of training samples

To understand the lower limit of training samples that SII-NowNet requires to maintain acceptable performance, the 1 hour lead time version was re-trained 6 times (for both initiation and intensification), each time halving the number of training samples that were used. Figure

12 shows the AUC scores and fraction of original training time (58 minutes and 3 seconds) for each halving iteration. For intensification a 94% reduction in the number of training samples, from a year of data to approximately three weeks of data (8,661 to 541), results in an 11% reduction in SII-NowNet's AUC score and a 99% reduction in training time. However, further reduction of training samples from 541 to 270 results in a rapid drop in AUC score from 0.84 to 0.5. This provides an approximate limit of training samples for achieving acceptable SII-NowNet intensification performance with two orders of magnitude greater speed. For initiation, SII-NowNet reduces more steadily as the number of samples reduces. A 75% reduction in the number of training samples, from a year of data to approximately three months of data (8,661 to 2,165), results in an 11% reduction in SII-NowNet's AUC score and a 75% reduction in training time. At this point SII-NowNet's AUC score is approximately equal to the initiation climatology (Figure 10b), providing an approximate limit of training samples for initiation.

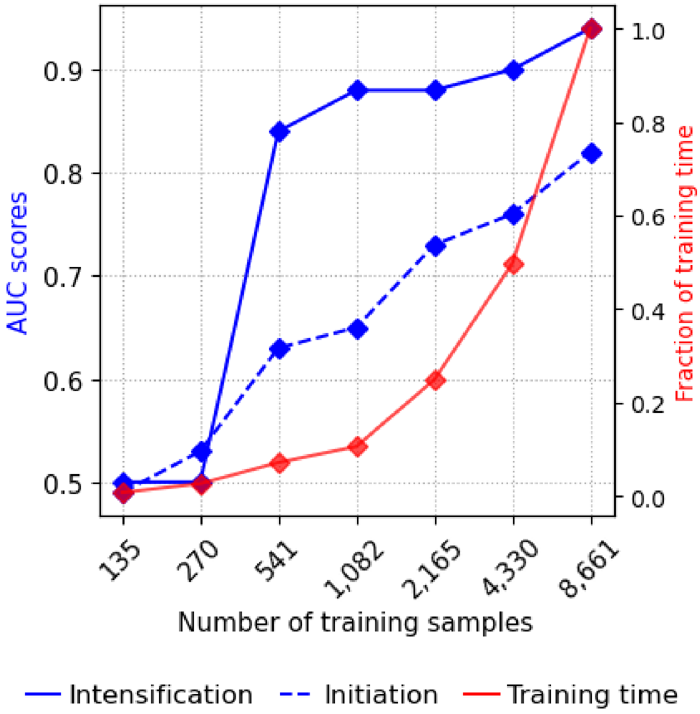

Fig 12. *Left axis* shows the AUC skill of 1 hour lead time SII-NowNet initiation (blue dashed) and intensification (blue solid) nowcasts as the number of training samples are continuously halved from the 8,661 (a year of data) to 135 (approximately three weeks). *Right axis* shows the fraction of time it took to train the model with this number of training samples, compared to the full training dataset (58 minutes and 3 seconds).

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

d) Ablation study

To gain insight into SII-NowNet's decision-making process, a permutation test is used to provide information on the relative influence that each of its inputs has on its predictions. Figure 13 shows the relative reduction in AUC score for SII-NowNet (for lead times that are within the limit of capability only) after each input predictor (BT T-0 and BT T-1) is, in turn, replaced with a Gaussian noise field. A greater relative reduction in AUC is associated with the predictor having a greater influence on the prediction (i.e. greater importance). In general, as lead time increases (and SII-NowNet skill decreases), there is a reduction in the influence of both predictors on SII-NowNet's nowcasts of initiation and intensification. As lead time increases, the BT data becomes less relevant for convective activity at the nowcast valid time, hence the predictors' influence on SII-NowNet reduces. The BT T-0 (the latest BT image) input produces the greatest reduction in AUC for SII-NowNet across all lead times. This is to be expected as it contains the most up-to-date information on cloud activity over the region. However, across all lead times, the BT T-1 permutation produces a minimum of 20% reduction in AUC score. This significant reduction suggests that, although not as important as the latest BT image, information in BT data from the previous hour is still useful for predicting the likelihood of convection development for the next few hours.

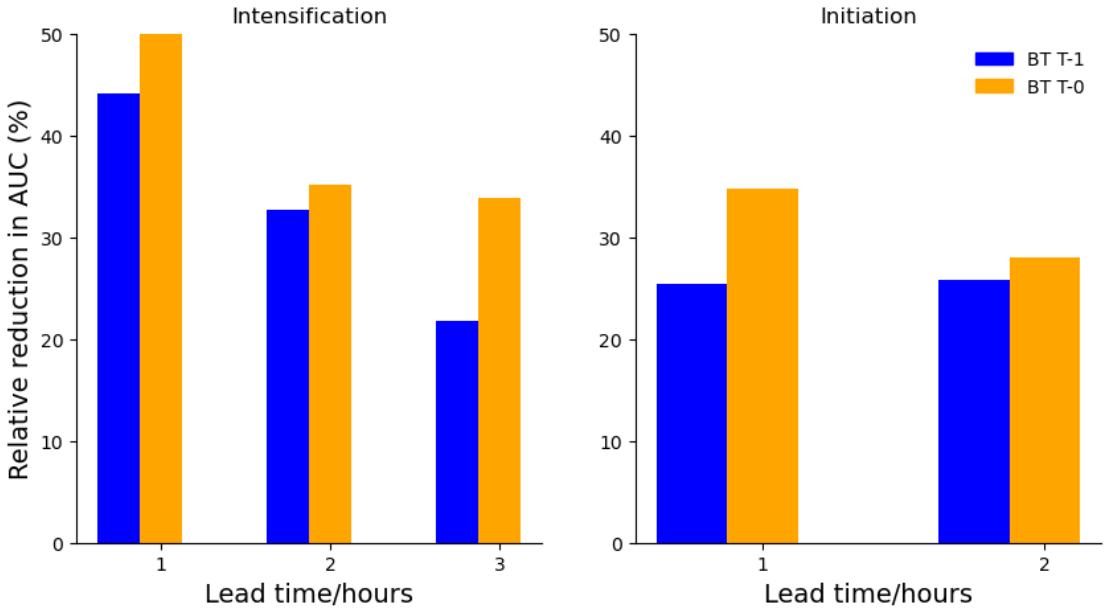

Fig 13. The mean relative reduction in the AUC score of SII-NowNet when each input channel (BT T-0 and BT T-1) is, in turn, replaced with a Gaussian noise field. The Gaussian noise replacement was repeated 20 times for each variable across each lead time that is considered skilful.

e) SII-NowNet applied across the Tropics

To provide evidence of SII-NowNet's generalisability to other Tropical regions, its performance is evaluated over four other domains in the Tropics (without any re-training or region-specific tuning): New Guinea, Zambia, Congo and West Africa.

Figure 14 presents a case study of SII-NowNet applied to Zambia (in the same format as in Figure 9). The intensification and initiation nowcasts are valid at 13:15 LT on 5th March 2024 and only lead times within SII-NowNet's limit of capability are presented (see Figure 10). Figures 14a and b show regions of cold cloud tops deepening over the east of the domain along with new initiations occurring ahead of these clouds towards the north of Zambia. Without any prior knowledge of this new domain, SII-NowNet appears to show skill at capturing these widespread events. At 1 hours' lead time, the high intensification nowcast probabilities (Figure 14d) align with the bulk of observed events over the eastern and northern regions of Zambia. Similarly, for the initiation nowcast (Figure 14k), SII-NowNet has captured a high proportion of the observed events e.g. over the central part of the domain. However, the initiation nowcast does not capture the full extent of the initiations over the northwestern part of the domain. At 2 hours' lead time the intensification (Figure 14e) and initiation (Figure 14i) nowcasts produce more blurred images with less precision, with the initiation nowcast missing more events in the northwest. However, the distribution of high probabilities still broadly captures a high proportion of events over the eastern part of the domain i.e. SII-NowNet continues to capture both events at 2 hours' lead time. For the 3 hour lead time intensification nowcast (Figure 14f), the high probabilities still align with some of the observed events in the east, however, reduced skill at this lead time results in larger areas of observed events being missed over the north of Zambia.

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

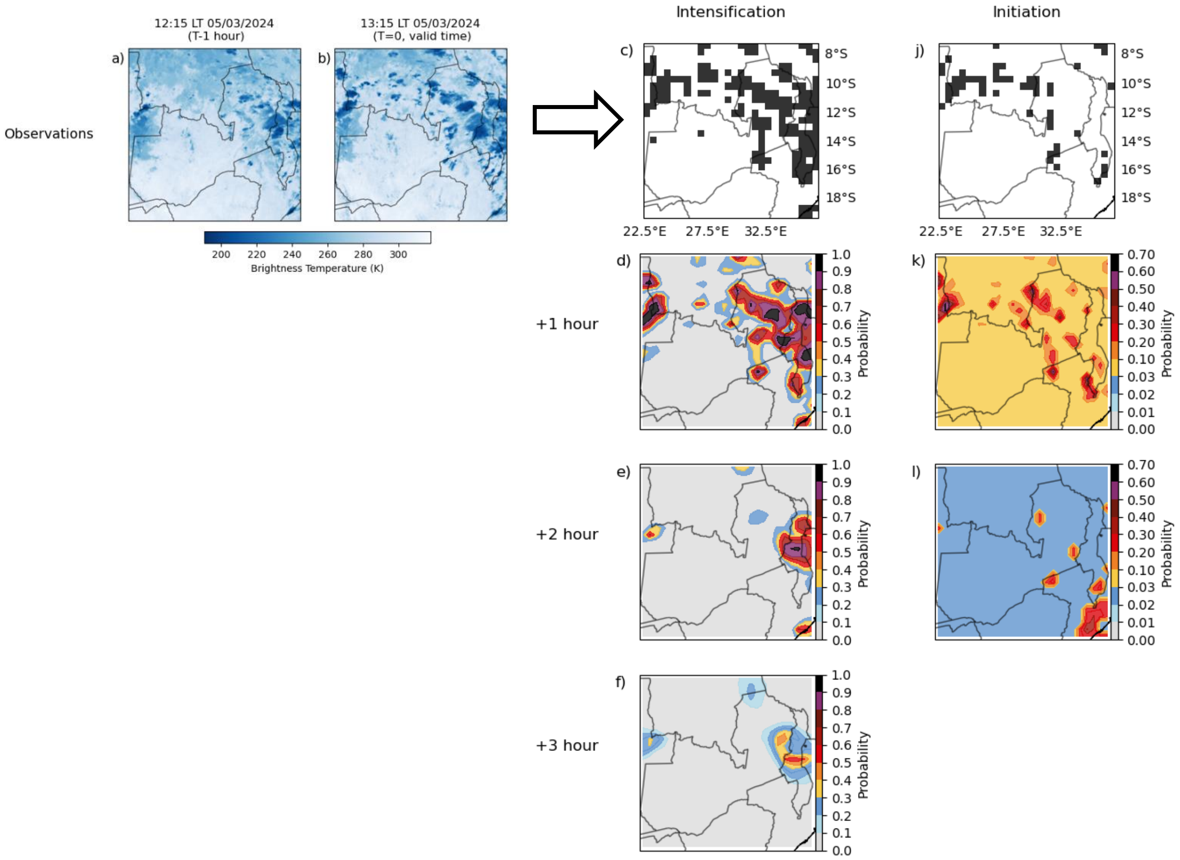

Fig 14. A case study for Zambia valid at 1315 LT on 15th March 2024. Presented in the same format as Figure 9 except only using SII-NowNet's lead times that are within its limit of capability.

Figure 15 shows the AUC scores for Zambia, New Guinea, Congo and West Africa when tested on 1 month of data (see section 2.a for details), along with the AUC scores over Sumatra for reference (see Figure 10). For each domain's AUC score, the corresponding climatology (horizontal dashed line) has been added for comparison. Although SII-NowNet has no knowledge of these new domains, it still produces comparable AUC scores to those over the Sumatran domain. At 1 hours' lead time, New Guinea produces the highest AUC score for both initiation and intensification. Comparing SII-NowNet's 1 hour lead time AUC scores over Sumatra to those over New Guinea shows there is only a 1% and 7% reduction for intensification and initiation, respectively. Over Africa, the West Africa domain shows the greatest improvement over climatology (21% performance gain for 1 hour intensification nowcasts). However, across all shown lead times for intensification and initiation, SII-NowNet outperforms the climatology baseline for every domain. Furthermore, although not shown here, each region also produces similar performance gain trends over persistence as shown in Figure 11.

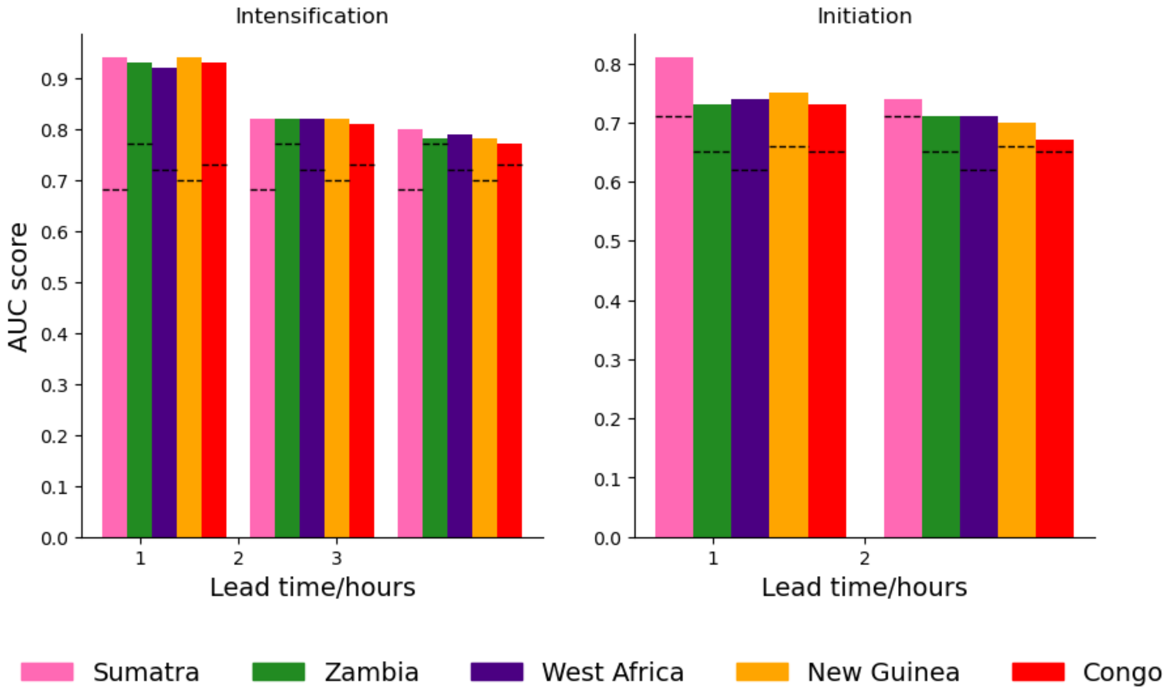

Fig 15. AUC scores for SII-NowNet when applied to four new domains across the Tropics (without any re-training or region-specific tuning), along with the AUC scores for Sumatra (taken from Figure 10). The climatology for each region is shown as a horizontal line for comparison. Only lead times within SII-NowNet's limit of capability are shown.

4. Conclusions and discussion

This study introduces SII-NowNet, a deep learning model that produces individual nowcasts of convection intensification and convection initiation in the Tropics. The model is driven by two freely available inputs: the latest BT image and the BT image from 1 hour earlier. This simplicity makes it advantageous as a nowcasting tool, as it can be quickly and easily implemented anywhere in the Tropics, which is covered by geostationary satellite brightness temperature observations. This study predominantly focuses on the application of SII-NowNet over Sumatra, Indonesia, evaluating nowcasts at 1 to 6 hour lead times. However, performance is also evaluated over New Guinea, Zambia, Congo and West Africa as evidence of its generalisability to other Tropical regions.

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

SII-NowNet performs best at nowcasting convection intensification. Over Sumatra, it outperforms STEPS, persistence and climatology across 1 – 6 hours' lead time, producing consistently higher likelihoods of capturing observed intensification events. SII-NowNet's nowcasts show reliability for up to approximately 3 hours' lead time, beyond which it begins to over-predict the number of intensification events (producing false alarms) i.e. SII-NowNet's intensification nowcasts have a 3 hour lead time limit of capability.

SII-NowNet also shows capability in predicting convection initiation events. SII-NowNet has learnt to associate the less frequent initiation events with greater uncertainty (compared to intensification events), resulting in it producing nowcasts with lower probabilities. Over Sumatra, it produces reliable nowcasts that outperform STEPS and climatology for 2 hours' lead time. However, beyond this lead time skill dramatically drops for SII-NowNet's initiation nowcasts, which lose reliability by producing a high proportion of false alarms.

SII-NowNet's results over Sumatra show that, by learning key patterns in the atmospheric conditions, it is able to discriminate between convective and non-convective events, proving its potential to support early weather warnings with multiple hours' lead time.

To the authors' knowledge there are no other tools that produce probabilistic nowcasts (on 1 – 6 hour lead times) of convection initiation and convection intensification in the Tropics, using solely satellite data. It is therefore not appropriate to make a direct comparison to other studies, but instead use them as context for model performance. In Korea (subtropics), previous studies have presented nowcasting tools that produce an 82% (Han et al., 2019) and 85% (Lee et al., 2017) chance of capturing observed convective events (both studies used radar data for identifying convection and produced deterministic models). For intensification events at a 1 hour lead time, SII-NowNet was thresholded to produce deterministic output with an 82% chance of capturing observed events. Therefore, not only does SII-NowNet provide probabilistic information, but it can also produce similar deterministic results to previous studies. Mecikalski et al., (2021) used radar, satellite and lightning data to produce a probabilistic convection initiation detection model (i.e. 0 hour lead time), which produced an AUC score (where higher is better) of 0.81, comparable to the AUC score of 0.82 produced by 1 hour lead time SII-NowNet initiation nowcasts. Lagerquist et al., (2021) used U-Net models (trained on radar data) to produce probabilistic fields of convection over Taiwan and showed similar reliability results for 1 hour lead time nowcasts. However, for 2 hours' lead time, their results show that the U-Net over-predicted the likelihood of convection (i.e. dropped in

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

reliability), whereas SII-NowNet maintains its reliability for both intensification and initiation nowcasts at this lead time.

A feature importance test of SII-NowNet's inputs showed that, as expected, the latest BT image has the most influence on its decision-making. However, the BT image from one hour earlier also demonstrated significant influence on SII-NowNet's intensification decisions across 1 – 3 hours' lead time. This suggests that SII-NowNet has learned that changes in the atmospheric conditions over the past hour provide relevant information for future convection development – a useful result for further development on convection prediction tools.

An investigation into how SII-NowNet's skill varies with the number of training samples showed that the model needed approximately three weeks of hourly training data to retain an acceptable level of intensification nowcast performance, resulting with 99% less training time. For initiation nowcasts, SII-NowNet requires approximately three months of hourly training data to retain acceptable skill, resulting in 75% less training time. These results have practical implications for future users of SII-NowNet when re-training the model in new regions. This low volume data requirement makes SII-NowNet a favourable tool for less-developed regions of the Tropics, whose national weather services may be limited by computational resource.

To highlight SII-NowNet's applicability to other regions of the Tropics, its performance was also evaluated over New Guinea, Zambia, Congo and West Africa. Without any prior knowledge of the new domain (i.e. no re-training or region-specific tuning), SII-NowNet was able to produce skilful nowcasts of intensification and initiation for these domains, with comparable skill scores to those produced over Sumatra. Overall, these results provide evidence that SII-NowNet can be applied to other areas of the Tropics and still produce skilful nowcasts. Future work will continue to test SII-NowNet's performance over these new domains by incorporating additional location-specific training.

In order to overcome the issue of sparse ground-based observations in the Tropics, this study uses cloud top brightness temperatures as a proxy for convection (which has known limitations e.g. contamination of anvil clouds). This means that SII-NowNet's nowcasts are unable to provide information on rainfall amounts but instead provide areas of warning/concern for deepening convection. However, deriving rainfall rates from satellite data is an advancing area of nowcasting (Amell et al., 2025), which provides an excellent opportunity for incorporation into SII-NowNet as a focus for continued development.

Another key limitation of this study is the methodologies used for identifying intensification and initiation events, which were shown to be highly sensitive to their pre-defined thresholds. Furthermore, storms that are growing/propagating at speeds ≥11 m/s may by mistakenly identified as an initiation. Nonetheless, the results show that, when provided with a set of thresholds, SII-NowNet has predictive capability and potential value for predicting convection development. Therefore, to overcome this limitation, future work will focus on developing and using more accurate and reliable datasets of convective events to improve the capability and robustness of SII-NowNet. Fan et al., (2024) used a ≥35 dBZ threshold across three different criteria to identify convection initiation pixels in radar reflectivity. They then applied a tracking algorithm (Lakshmanan and Smith, 2010) to filter out the initiations that were related to pre-existing convection. Incorporating this storm tracking technique into the identification method in this study could be tested for improving SII-NowNet.

Future improvements to SII-NowNet could also be achieved through additional input variables. For the practical purposes of implementation (in resource constrained settings), SII-NowNet only requires two BT input fields. However, increasing the temporal resolution of the input BT fields may allow SII-NowNet to better capture the atmospheric dynamics. Alternatively, the Global Forecast System (National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, 2015) produces live NWP output, which provides a range of atmospheric variables that could be incorporated and tested as SII-NowNet inputs.

Overall, SII-NowNet has shown that it can provide skilful probabilistic nowcasts of Tropical convection development in satellite observations. A key advantage of SII-NowNet is in its use of simple, freely available inputs, meaning that it can be applied to any region that is covered by a geostationary satellite. This versatility, along with the planned future developments, mean that SII-NowNet has great potential for future use as an operational nowcasting tool. Currently Badan Meteorologi, Klimatologi, dan Geofisika (BMKG; Indonesian national weather service) have adopted SII-NowNet and are running it in an operational trial, which can be seen on their website (https://web-meteo.bmkg.go.id/id/peringatan/sii-nownet).

Acknowledgments. The authors would like to thank the NERC SENSE CDT (NE/T00939X/1) who provided the core funding for this work. Birch and Marsham were additionally funded by the NERC TerraMaris project (NE/R016739/1). Thanks to NEODASS, who provided the GPU facilities for training machine learning models and technical guidance. Thanks to the reviewers, who helped strengthen the study and improve its robustness.

Data Availability Statement. The Githhub page containing all the code for developing SII-NowNet is available at https://github.com/eejasm/SII-NowNet/tree/main. Only freely available brightness temperature data was used in this work, which can be accessed for Sumatra and New Guinea at https://www.icare.univ-lille.fr/asd-content/archive/?dir=GEO/HIMAWARI%201407/GEO_L1B.v1.06/ and for Zambia, Congo, West Africa at https://user.eumetsat.int/catalogue/EO:EUM:DAT:MSG:HRSEVIRI/access.

REFERENCES

Amell, A., Hee, L., Pfreundschuh, S., Eriksson, P., 2025. Probabilistic near real-time retrievals of Rain over Africa using deep learning. https://doi.org/10.22541/essoar.173867530.07619555/v1

Ayzel, G., Scheffer, T., Heistermann, M., 2020. RainNet v1.0: a convolutional neural network for radar-based precipitation nowcasting. Geosci. Model Dev. 13, 2631–2644. https://doi.org/10.5194/gmd-13-2631-2020

Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., Kumagai, Y., Miyakawa, T., Murata, H., Ohno, T., Okuyama, A., Oyama, R., Sasaki, Y., Shimazu, Y., Shimoji, K., Sumida, Y., Suzuki, M., Taniguchi, H., Tsuchiyama, H., Uesawa, D., Yokota, H., Yoshida, R., 2016. An Introduction to Himawari-8/9— Japan’s New-Generation Geostationary Meteorological Satellites. Journal of the Meteorological Society of Japan 94, 151–183. https://doi.org/10.2151/jmsj.2016-009

Bowler, N.E., Pierce, C.E., Seed, A.W., 2006. STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. Q. J. R. Meteorol. Soc. 132, 2127–2155. https://doi.org/10.1256/qj.04.100

Browning, K.A., 1980. Review Lecture: Local weather forecasting. Proc. R. Soc. Lond. A 371, 179–211. https://doi.org/10.1098/rspa.1980.0076

Burton, R.R., Blyth, A.M., Cui, Z., Groves, J., Lamptey, B.L., Fletcher, J.K., Marsham, J.H., Parker, D.J., Roberts, A., 2022. Satellite-based nowcasting of West African mesoscale storms has skill at up to four hours lead time. Weather and Forecasting. https://doi.org/10.1175/WAF-D-21-0051.1

Crook, J., Morris, F., Fitzpatrick, R.G.J., Peatman, S.C., Schwendike, J., Stein, T.H., Birch, C.E., Hardy, S., Yang, G., 2024. Impact of the Madden–Julian oscillation and equatorial waves on tracked mesoscale convective systems over southeast Asia. Quart J Royal Meteoro Soc 150, 1724–1751. https://doi.org/10.1002/qj.4667

Fan, D., Greybush, S.J., Clothiaux, E.E., Gagne, D.J., 2024. Physically Explainable Deep Learning for Convective Initiation Nowcasting Using GOES-16 Satellite Observations. Artificial Intelligence for the Earth Systems. https://doi.org/10.1175/AIES-D-23-0098.1

Feng, Z., Leung, L.R., Liu, N., Wang, J., Houze, R.A., Li, J., Hardin, J.C., Chen, D., Guo, J., 2021. A Global High-Resolution Mesoscale Convective System Database Using Satellite-Derived Cloud Tops, Surface Precipitation, and Tracking. JGR Atmospheres 126, e2020JD034202. https://doi.org/10.1029/2020JD034202

Ferrett, S., Frame, T.H.A., Methven, J., Holloway, C.E., Webster, S., Stein, T.H.M., Cafaro, C., 2021. Evaluating convection-permitting ensemble forecasts of precipitation over Southeast Asia. Weather and Forecasting. https://doi.org/10.1175/WAF-D-20-0216.1

Fletcher, J.K., Diop, C.A., Adefisan, E., Ahiataku, M.A., Ansah, S.O., Birch, C.E., Burns, H.L., Clarke, S.J., Gacheru, J., James, T.D., Tuikong, C.K.N., Koros, D., Indasi, V.S., Lamptey, B.L., Lawal, K.A., Parker, D.J., Roberts, A.J., Stein, T.H.M., Visman, E., Warner, J., Woodhams, B.J., Youds, L.H., Ajayi, V.O., Bosire, E.N., Cafaro, C., Camara, C.A.T., Chanzu, B., Dione, C., Gitau, W., Groves, D., Groves, J., Hill, P.G., Ishiyaku, I., Klein, C.M., Marsham, J.H., Mutai, B.K., Ndiaye, P.N., Osei, M., Popoola, T.I., Talib, J., Taylor, C.M., Walker, D., 2023. Tropical Africa’s First Testbed for High-Impact Weather Forecasting and Nowcasting. Bulletin of the American Meteorological Society 104, E1409–E1425. https://doi.org/10.1175/BAMS-D-21-0156.1

Germann, U., Zawadzki, I., 2002. Scale-Dependence of the Predictability of Precipitation from Continental Radar Images. Part I: Description of the Methodology. Mon. Wea. Rev. 130, 2859–2873. https://doi.org/10.1175/1520-0493(2002)130<2859:SDOTPO>2.0.CO;2

Haiden, T., Kann, A., Wittmann, C., Pistotnik, G., Bica, B., Gruber, C., 2011. The Integrated Nowcasting through Comprehensive Analysis (INCA) System and Its Validation over the Eastern Alpine Region. Weather and Forecasting 26, 166–183. https://doi.org/10.1175/2010WAF2222451.1

Han, D., Lee, J., Im, J., Sim, S., Lee, S., Han, H., 2019. A Novel Framework of Detecting Convective Initiation Combining Automated Sampling, Machine Learning, and Repeated Model Tuning from Geostationary Satellite Data. Remote Sensing 11, 1454. https://doi.org/10.3390/rs11121454

Han, L., Sun, J., Zhang, W., Xiu, Y., Feng, H., Lin, Y., 2017. A machine learning nowcasting method based on real-time reanalysis data. JGR Atmospheres 122, 4038–4051. https://doi.org/10.1002/2016JD025783

Hill, P.G., Stein, T.H.M., Roberts, A.J., Fletcher, J.K., Marsham, J.H., Groves, J., 2020. How skilful are NOWCASTING SATELLITE APPLICATIONS FACILITY products for tropical Africa? Meteorol Appl 27. https://doi.org/10.1002/met.1966

Horn, B.K.P., Schunck, B.G., 1981. Determining optical flow. Artificial Intelligence 17, 185–203. https://doi.org/10.1016/0004-3702(81)90024-2

Lagerquist, R., Stewart, J., Ebert-Uphoff, I., Christina, K., 2021. Using Deep Learning to Nowcast the Spatial Coverage of Convection from Himawari-8 Satellite Data. Monthly Weather Review 149, 3897–3921. https://doi.org/10.1175/MWR-D-21-0096.1

Lakshmanan, V., Smith, T., 2010. An Objective Method of Evaluating and Devising Storm-Tracking Algorithms. Weather and Forecasting 25, 701–709. https://doi.org/10.1175/2009WAF2222330.1

Lee, S., Han, H., Im, J., Jang, E., Lee, M.-I., 2017. Detection of deterministic and probabilistic convection initiation using Himawari-8 Advanced Himawari Imager data. Atmos. Meas. Tech. 10, 1859–1874. https://doi.org/10.5194/amt-10-1859-2017

Line, W.E., Schmit, T.J., Lindsey, D.T., Goodman, S.J., 2016. Use of Geostationary Super Rapid Scan Satellite Imagery by the Storm Prediction Center*. Weather and Forecasting 31, 483–494. https://doi.org/10.1175/WAF-D-15-0135.1

Lo, J.C., Orton, T., 2016. The general features of tropical Sumatra Squalls. Weather 71, 175–178. https://doi.org/10.1002/wea.2748

Lo, J.C.-F., Orton, T., 2016. The general features of tropical Sumatra Squalls 71, 4.

Machado, L.A.T., Laurent, H., 2004. The Convective System Area Expansion over Amazonia and Its Relationships with Convective System Life Duration and High-Level Wind Divergence. Mon. Wea. Rev. 132, 714–725. https://doi.org/10.1175/1520-0493(2004)132<0714:TCSAEO>2.0.CO;2

Marcos, 2015. NWC SAF convective precipitation product from MSG: A new day-time method based on cloud top physical properties. Tethys. https://doi.org/10.3369/tethys.2015.12.01

McGovern, A., Chase, R.J., Flora, M., Gagne, D.J., Lagerquist, R., Potvin, C.K., Snook, N., Loken, E., 2023. A Review of Machine Learning for Convective Weather. Artificial Intelligence for the Earth Systems 2, e220077. https://doi.org/10.1175/AIES-D-22-0077.1

Mecikalski, J.R., Sandmæl, T.N., Murillo, E.M., Homeyer, C.R., Bedka, K.M., Apke, J.M., Jewett, C.P., 2021. Random Forest Model to Assess Predictor Importance and Nowcast Severe Storms using High-Resolution Radar–GOES Satellite–Lightning Observations. Monthly Weather Review. https://doi.org/10.1175/MWR-D-19-0274.1

Mecikalski, J.R., Williams, J.K., Jewett, C.P., Ahijevych, D., LeRoy, A., Walker, J.R., 2015. Probabilistic 0–1-h Convective Initiation Nowcasts that Combine Geostationary Satellite Observations and Numerical Weather Prediction Model Data. Journal of Applied Meteorology and Climatology 54, 1039–1059. https://doi.org/10.1175/JAMC-D-14-0129.1

Mori, S., Jun-Ichi, H., Tauhid, Y.I., Yamanaka, M.D., Okamoto, N., Murata, F., Sakurai, N., Hashiguchi, H., Sribimawati, T., 2004. Diurnal Land–Sea Rainfall Peak Migration over Sumatera Island, Indonesian Maritime Continent, Observed by TRMM Satellite and Intensive Rawinsonde Soundings. Mon. Wea. Rev. 132, 2021–2039. https://doi.org/10.1175/1520-0493(2004)132<2021:DLRPMO>2.0.CO;2

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, 2015. NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive. https://doi.org/10.5065/D65D8PWK

Peatman, S.C., Birch, C.E., Schwendike, J., Marsham, J.H., Dearden, C., Webster, S., Neely, R.R., Matthews, A.J., 2023. The Role of Density Currents and Gravity Waves in the Offshore Propagation of Convection over Sumatra. Monthly Weather Review 151, 1757–1777. https://doi.org/10.1175/MWR-D-22-0322.1

Permana, D.S., Hutapea, T.D., Praja, A.S., Paski, J.A.I., Makmur, E.E.S., Haryoko, U., Umam, I.H., Saepudin, M., Adriyanto, R., 2019. The Indonesia In-House Radar Integration System (InaRAISE) of Indonesian Agency for Meteorology Climatology and Geophysics (BMKG): Development, Constraint, and Progress. IOP Conf. Ser.: Earth Environ. Sci. 303, 012051. https://doi.org/10.1088/1755-1315/303/1/012051

Porson, A.N., Hagelin, S., Boyd, D.F.A., Roberts, N.M., North, R., Webster, S., Lo, J.C., 2019. Extreme rainfall sensitivity in convective-scale ensemble modelling over Singapore. Q.J.R. Meteorol. Soc. 145, 3004–3022. https://doi.org/10.1002/qj.3601

Pulkkinen, S., Nerini, D., Pérez Hortal, A.A., Velasco-Forero, C., Seed, A., Germann, U., Foresti, L., 2019. Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1.0). Geosci. Model Dev. 12, 4185–4219. https://doi.org/10.5194/gmd-12-4185-2019

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S., Prudden, R., Mandhane, A., Clark, A., Brock, A., Simonyan, K., Hadsell, R., Robinson, N., Clancy, E., Arribas, A., Mohamed, S., 2021. Skilful precipitation nowcasting using deep generative models of radar. Nature 597, 672–677. https://doi.org/10.1038/s41586-021-03854-z

Roberts, A.J., Fletcher, J.K., Groves, J., Marsham, J.H., Parker, D.J., Blyth, A.M., Adefisan, E.A., Ajayi, V.O., Barrette, R., de Coning, E., Dione, C., Diop, A., Foamouhoue, A.K., Gijben, M., Hill, P.G., Lawal, K.A., Mutemi, J., Padi, M., Popoola, T.I., Rípodas, P., Stein, T.H.M., Woodhams, B.J., 2022. Nowcasting for AFRICA : advances, potential and value. Weather 77, 250–256. https://doi.org/10.1002/wea.3936

Roca, R., Fiolleau, T., Bouniol, D., 2017. A Simple Model of the Life Cycle of Mesoscale Convective Systems Cloud Shield in the Tropics. Journal of Climate 30, 4283–4298. https://doi.org/10.1175/JCLI-D-16-0556.1

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. https://doi.org/10.48550/ARXIV.1505.04597

Rouault, E., Warmerdam, F., Schwehr, K., Kiselev, A., Butler, H., Łoskot, M., Szekeres, T., Tourigny, E., Landa, M., Miara, I., Elliston, B., Chaitanya, K., Plesea, L., Morissette, D., Jolma, A., Dawson, N., Baston, D., de Stigter, C., Miura, H., 2023. GDAL. https://doi.org/10.5281/ZENODO.5884351

Saltikoff, E., Friedrich, K., Soderholm, J., Lengfeld, K., Nelson, B., Becker, A., Hollmann, R., Urban, B., Heistermann, M., Tassone, C., 2019. An Overview of Using Weather Radar for Climatological Studies: Successes, Challenges, and Potential. Bulletin of the American Meteorological Society 100, 1739–1752. https://doi.org/10.1175/BAMS-D-18-0166.1

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., Ratier, A., 2002. Supplement to An Introduction to Meteosat Second Generation (MSG): SEVIRI CALIBRATION. Bull. Amer. Meteor. Soc. 83, 992–992. https://doi.org/10.1175/BAMS-83-7-Schmetz-2

Schumacher, R.S., Rasmussen, K.L., 2020. The formation, character and changing nature of mesoscale convective systems. Nat Rev Earth Environ 1, 300–314. https://doi.org/10.1038/s43017-020-0057-7

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W., Woo, W., 2017. Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model. https://doi.org/10.48550/ARXIV.1706.03458

Smith, J., Birch, C., Marsham, J., Peatman, S., Bollasina, M., Pankiewicz, G., 2024. Evaluating pySTEPS optical flow algorithms for convection nowcasting over the Maritime Continent using satellite data. Nat. Hazards Earth Syst. Sci. 24, 567–582. https://doi.org/10.5194/nhess-24-567-2024

Vogel, P., Knippertz, P., Fink, A.H., Schlueter, A., Gneiting, T., 2020. Skill of Global Raw and Postprocessed Ensemble Predictions of Rainfall in the Tropics. Weather and Forecasting 35, 2367–2385. https://doi.org/10.1175/WAF-D-20-0082.1

Wang, R., Fung, J.C.H., Lau, A.K.H., 2024. Skillful Precipitation Nowcasting Using Physical-Driven Diffusion Networks. Geophysical Research Letters 51, e2024GL110832. https://doi.org/10.1029/2024GL110832

World Meteorological Organization [WWW Document], 2023. . Early Warnings For All Initiative scaled up into action on the ground. URL https://public.wmo.int/en/media/press-release/early-warnings-all-initiative-scaled-action-ground

Yang, G.-Y., Slingo, J., 2001. The Diurnal Cycle in the Tropics. Mon. Wea. Rev. 129, 784–801. https://doi.org/10.1175/1520-0493(2001)129<0784:TDCITT>2.0.CO;2

Youds, L., Parker, D., Adefisan, E.A., Antwi-Agyei, P., Bain, C., Black, E., Blyth, A., Dougill, A., Hirons, L., Indasi, V., Lamptey, B., Marshall, F., Marsham, J., Stein, T., Taylor, C., todd, martin, Visman, E., Woolnough, S., 2021. GCRF African SWIFT and ForPAc SHEAR White Paper on the Potential of Operational Weather Prediction to Save Lives and Improve Livelihoods and Economies in Sub-Saharan Africa. White Rose University Consortium. https://doi.org/10.5518/100/79

Zhang, Y., Long, M., Chen, K., Xing, L., Jin, R., Jordan, M.I., Wang, J., 2023. Skilful nowcasting of extreme precipitation with NowcastNet. Nature 619, 526–532. https://doi.org/10.1038/s41586-023-06184-4

Accepted for publication in *Artificial Intelligence for the Earth Systems.* DOI 10.1175/AIES-D-25-0043.1.