

RESEARCH ARTICLE

Nowcasting convective cores using deep learning

Jawairia A. Ahmad¹  | Cornelia Klein¹  | Christopher Taylor^{1,2}  |
Seonaid R. Anderson¹  | Steven C. Wells¹  | David C. Hogg³ 

¹UK Centre for Ecology and Hydrology,
Wallingford, UK

²National Centre for Earth Observation,
Leicester, UK

³School of Computer Science, University
of Leeds, Leeds, UK

Correspondence

Jawairia A. Ahmad, UK Centre for
Ecology and Hydrology, Crowmarsh
Gifford, Wallingford, OX10 8BB, UK.
Email: jawahm@ceh.ac.uk

Funding information

Natural Environment Research Council,
Grant/Award Number: NE/Y000420/1

Abstract

Extreme precipitation events in the Tropics are often linked to convective storms. Storm nowcasting systems for early warning have been recognized as a necessary adaptation strategy to increase public resilience in a rapidly evolving climate. Ongoing advances in deep-learning techniques now allow better exploitation of satellite data, including where ground-based radar networks and computational resources are limited. This study presents an easily reproducible model (Network for Nowcasting Convective Cores, NetNCC) for nowcasting convective cores in near real time at a fine spatial resolution (~ 3 km). NetNCC uses Meteosat Second Generation thermal infrared brightness temperature images at three successive time steps to predict the probability of convective cores for lead times between 1 and 6 hr. A spatially enhanced loss function (fractions skill score) captures the atmospheric conditions within the neighbourhood of each pixel. Skilful probabilistic predictions are achieved at fine spatial resolutions (< 30 km) for short lead times of < 2 hr and at coarser resolution (180–270 km) for lead times ≥ 4 hr. Nowcast accuracy shows a clear diurnal cycle, with higher accuracy during the daytime when convection frequency is at a maximum in eastern and southern Africa. NetNCC performs better for short-lived convective cores that develop over eastern Africa than for cores embedded in typically long-lived West African convection. Compared with persistence, NetNCC nowcasts are considerably better particularly for fine spatial scales. Further input information on atmospheric and land-surface states alongside thermal infrared imagery may be required to improve longer (≥ 4 hr) lead-time predictions. Nevertheless, our results indicate considerable potential of NetNCC for nowcasting convective activity, and hence intense rainfall, in at-risk regions and illustrate the utility to forecasters for operational applications.

KEYWORDS

Africa, convective cloud cores, deep learning, nowcasting, U-Nets

1 | INTRODUCTION

Weather (in particular precipitation) forecasting several hours ahead has gained traction in recent years owing to increasing frequency and costs of impactful convective events as the Earth's climate is warming (Allen & Ingram, 2002; Westra *et al.*, 2014). Intensely precipitating convective storms in the Tropics can cause considerable damage, leading to flash floods and loss of lives and livelihoods (Maqtan *et al.*, 2022). Thus, short-range forecasting or “nowcasting” (also known as near-real-time prediction) is critical for effective early warning of severe weather and related natural hazards (Šakić Trogrlić *et al.*, 2022).

Contemporary methods of weather nowcasting include numerical weather prediction (NWP) models (Roberts *et al.*, 2018), probabilistic methods based on historical climatologies (Anderson *et al.*, 2023), statistical extrapolation of cloud or radar data (Imhoff *et al.*, 2020), and, increasingly, deep-learning-based nowcasting frameworks (Han *et al.*, 2019; Price *et al.*, 2025; Shi *et al.*, 2015). Based on the availability of satellite data and computational advances, nowcasting has evolved spatially as well as temporally. Higher-resolution numerical modelling and fine-scale satellite observations have significantly advanced the accuracy of rainfall nowcasting (Lam *et al.*, 2023). In addition to attempts at rainfall nowcasting where radar networks allow, indirect methods based on tracking storms or cold clouds have also helped to improve contemporary weather nowcasting efforts (Lakshmanan & Smith, 2010; Lehmann & Coumou, 2015).

Predicting the development of convective systems is an ongoing effort that is limited by uncertainties in initial conditions and the dependence of convective development on even small perturbations, error propagation within modelled forecasts, and limitations in model resolutions affecting the representation of convective systems (Zhao, 2014). This means that NWP models struggle to correctly spin up and position convective storms in space and time, even when provided with the most recent observational initial conditions. Short-term nowcasting strongly relies on satellite and, where available, in-situ observations that provide useful information in real time regarding convective system development (Li *et al.*, 2024). Overshooting top detection within convective clouds using infrared imagery can help improve the prediction of high-intensity storms (Ai *et al.*, 2017; Bedka & Khlopenkov, 2016). Hence, model-based systems that can indicate broader regions of new convective activity and observation-based systems that accurately track and extrapolate existing convection are both important in further advancing the prediction of convective storms.

Traditional observation-based nowcasting techniques, such as optical flow-based extrapolation of radar or cloud imagery implemented in tools such as Pysteps (Pulkkinen *et al.*, 2019) excel at capturing the advection of existing precipitation structures, which has been shown to have some skill over Africa (Burton *et al.*, 2022). Vector-flow extrapolation, however, struggles to predict storm initiation, intensification, or dissipation, which involve inherently nonlinear processes. On the other hand, simple statistical methods based on local conditional climatologies are able to include convective initiation and dissipation to some degree, based on typical historical convection life cycles including effects of landscape features such as coastlines and topography Anderson *et al.* (2023). However, this method is strongly limited by local sample sizes of convective events and sensitive to event pooling decisions. In contrast, deep-learning methods can learn complex spatio-temporally transferable relationships directly from historical data, offering the potential to recognize precursors to convective growth or decay with less dependency on local convection frequencies. Hence, advances in machine learning have opened promising avenues for improving short-range forecasts of convective storms.

Recent studies have established the utility of deep-learning-based methods for weather forecasting compared with traditional NWP models. In some cases, data-driven forecasting systems, such as GraphCast (Lam *et al.*, 2023) and Pangu-Weather (Bi *et al.*, 2023), have outperformed operational forecasting systems. The deep-learning architectures used in both GraphCast and Pangu-Weather are able to learn the mapping between different atmospheric variables to predict the short-term weather ahead. However, both these models follow a big data approach that is based on the idea of using any and all available information for model training. This makes them black-box methods with limited physical tractability and provides limited information about the geophysical processes that influence the weather at different scales and the relationships between the different atmospheric variables. In this context, machine-learning frameworks tailored for convective activity offer a complementary path: they can be trained specifically on satellite-observed signatures of convective evolution, run efficiently at high spatial resolution, and incorporate physical constraints through expert-led choice of input channels.

Convective systems have distinct spatial features that evolve with time as the systems mature. A number of drivers impact the evolution of convective systems, such as land-atmosphere interactions (Barton *et al.*, 2025) and wind shear (Alfaro & Khairoutdinov, 2015). This, in turn, results in inherent spatial features in convective systems that expand or disintegrate over time and

space. In the Tropics, convective systems have a distinct diurnal cycle as well that differs over ocean and land surface (Hendon & Woodberry, 1993). Deep convection over ocean is maximum during the early morning. Over land, convective activity is most intense between noon and evening, with local variations due to orographic lifting and land/sea breeze circulations (Yang & Slingo, 2001). Mesoscale convective systems, particularly in the Sahel region, are further impacted by soil moisture gradients (Klein & Taylor, 2020) such that dry soil conditions enhance propagating mesoscale convective systems. Spatio-temporal modelling of convective systems across different scales is therefore impacted by the complexity of interactions between the different drivers of convective activity.

In this study, we focus on convective cores that represent the deep convective part of a convective cloud system where high updraughts occur (Feng *et al.*, 2011). Different machine-learning (or deep-learning) architectures ranging from medium to highly complex have been proposed in the last couple of years focused on predicting precipitation (not convective cores), often using outputs from NWP models or reanalysis data (Bi *et al.*, 2023; Lam *et al.*, 2023). However, there is limited existing research on nowcasting convective cores. The rapid evolution of convective storms is well captured by convective cores. Since deep convection precedes convective rainfall, nowcasts of convective cores can provide a better lead time for predicting intense rainfall-induced hazardous weather events. In this study, we develop a nowcasting framework for deep convection (with a focus on Africa) that is easily run operationally and provides real-time nowcasts at short-range time periods for regional forecasters. The development of the framework was guided by four key requirements: (a) to provide near-real-time predictions, (b) to be physically tractable (i.e., changes in model output due to different physical processes could be tracked to some extent), (c) to be computationally flexible (i.e., easy to train and test for other regions), and (d) to be easily reproducible. These operational restrictions limit the type of data that could be used for training, the algorithm selection and complexity, and the grid scale and resolution. However, they provide a valuable opportunity to explore and narrow down which information types and sources are critical for accurate nowcasting at short range. It is common for deep-learning methods to be black-box models, such that very little physical meaning can be extracted from such models. Limiting the range of input data sources used in the model and testing the model performance across different climates helps to maintain some physical tractability. Further details about the model and its development are included in Section 2.

2 | METHODOLOGY

In satellite imagery, convective cores can be identified as very cold, often overshooting cloud tops (Bedka & Khlopenkov, 2016) and are frequently associated with extreme precipitation (Klein *et al.*, 2018). The separation between the deep convective part of the cloud and the stratiform anvil, associated with typically lighter precipitation, has traditionally used thresholds for cold cloud-top temperatures (Lin *et al.*, 2006; Machado *et al.*, 1998). Convective cores are important in storm monitoring as they provide a reasonable proxy for regions most likely to have intense precipitation (Feng *et al.*, 2011). Klein *et al.* (2018) used wavelet scale analysis to identify convective core boundaries using the gradient in thermal infrared (TIR) brightness temperatures T_b in mesoscale convective systems. For our model, we use this method to detect and subsequently predict convective cores across the study regions.

2.1 | Network for Nowcasting Convective Cores

The Network for Nowcasting Convective Cores (NetNCC) tool is based on a U-Net architecture Ronneberger *et al.* (2015) and predicts the propagation of convective cores by leveraging the spatio-temporal flexibility of convolutional neural networks (LeCun & Bengio, 1998). A U-Net architecture comprises convolutional neural network layers that successively reduce the dimension of the input information by extracting important features via pooling and then upscale to the dimension of the target image (Figure 1). The Tensorflow software library was used to code the algorithm in Python (Abadi *et al.*, 2015). U-Nets Ronneberger *et al.* (2015) have exhibited considerable potential in capturing spatial features in weather data, in particular precipitation, and propagating them over subsequent time periods (Kaparakis & Mehrkanoon, 2023; Trebing *et al.*, 2021). We carried out extensive test simulations to find the minimum number of parameters (and parameter values) required to achieve the lowest possible training loss value. The motivation behind this approach was to keep the model as simple and accurate as possible such that it could be easily used operationally in real time by forecasters. It also helped in keeping the computational memory required for training and testing low while also rendering it easy to test input sensitivities.

The model takes three images of Meteosat Second Generation (MSG) TIR T_b with an hourly interval as input to predict the probability of cores for 1, 2, 4 and 6 hr lead times. In addition to core propagation,

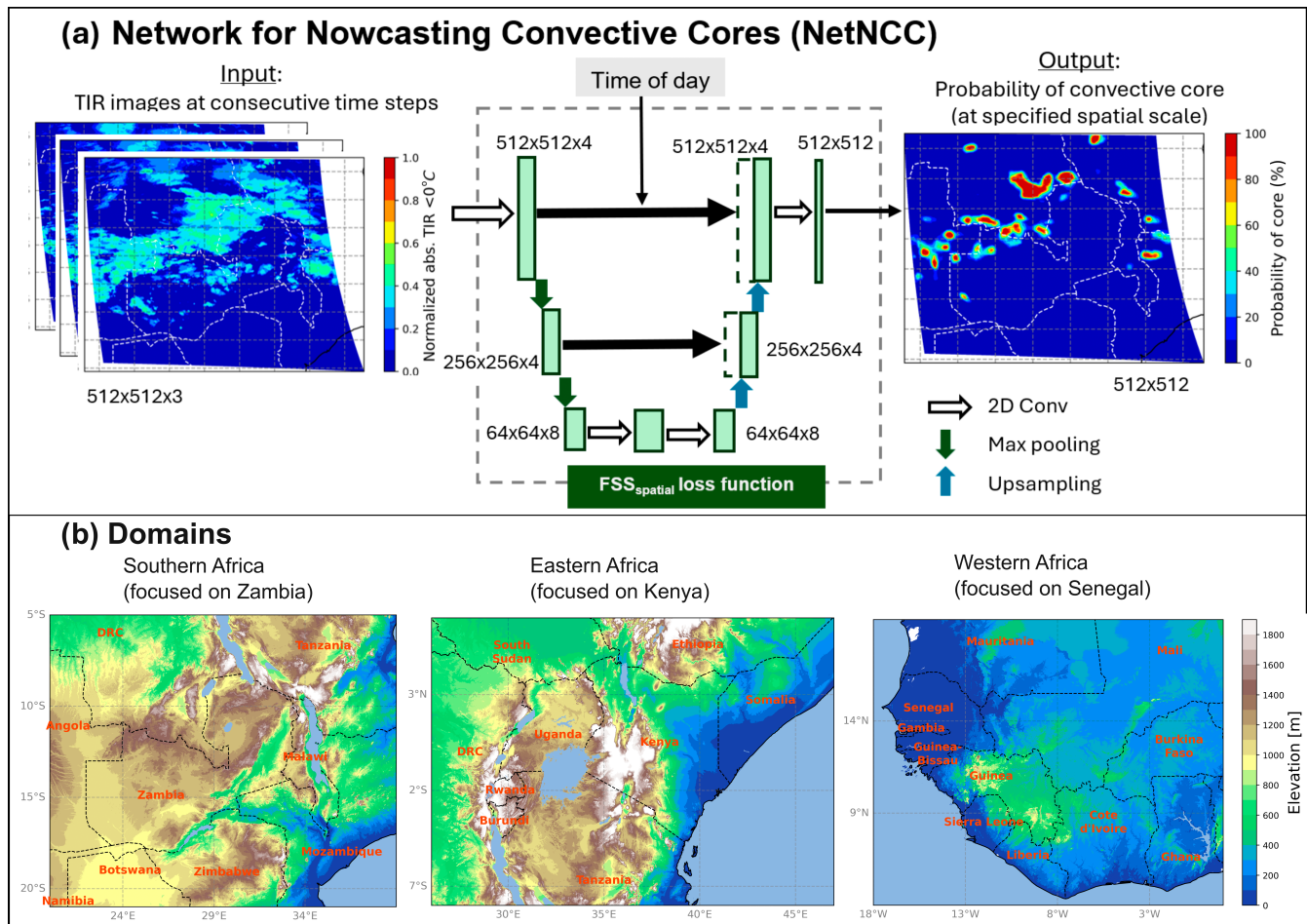


FIGURE 1 (a) NetNCC has a U-Net structure consisting of successive convolutional neural network layers. The numbers indicate the sizes of the different feature maps. (b) The topographic variation within the selected domains in Africa. The black dashed lines show country boundaries, the country names are in red, and the water bodies are in light blue colour. TIR: thermal infrared; FSS: fractions skill score.

NetNCC thus learns to detect convective cores directly from TIR imagery, allowing convective core initiations at lead times diagnosed from preceding cloud pattern changes. We varied the number of input images preceding t_0 (where t_0 is the time at which the prediction is initialized) from two to five and optimized other parameter values for each. Performance was found to be optimal with three images (t_0 , $t_0 - 1$ hr, and $t_0 - 2$ hr) as input. Increasing the number of inputs beyond three decreased the sensitivity of the predictions to the image at t_0 , thereby resulting in a decrease in prediction accuracy in some cases.

2.1.1 | Spatially enhanced loss function

Loss functions optimized for each pixel do not take into account the spatial structures in atmospheric data that define the spatio-temporal changes in interlinked state

variables. Selection of the appropriate loss function is important for deep-learning-based models. Optimizing the loss function for each pixel independently neglects the spatial correlation between neighbouring pixels. In other words, pixel-wise loss functions do not consider the occurrence of convection at nearby pixels and thus do not account for spatial features in the neighbourhood when computing the loss function. Lagerquist *et al.* (2021) overcame the lack of spatial information in training loss functions in contemporary deep-learning-based nowcasting models by using spatially enhanced loss functions. These are loss functions that use information from the pixel neighbourhood to highlight spatial features (convection in this case) when computing the loss during training. In their subsequent study, Lagerquist and Ebert-Uphoff (2022) used neighbourhood and spectral filters in the loss functions and provided a comprehensive analysis of using different scale thresholds and their impact on the predicted target (convection masks in their

case). Similar concepts have been used in other studies to include local spatial gradients during loss calculation (Eigen & Fergus, 2015; Heim & Avery, 2019; Stengel *et al.*, 2020).

In this study, we use the fractions skill score (FSS) with a neighbourhood filter as the loss function based on the comparatively better performance noted by Lagerquist and Ebert-Uphoff (2022). FSS compares the agreement of fractional coverage within a defined window size between prediction and observation and is a commonly used metric for verifying the skill of predictions (usually precipitation) across scales (Roberts & Lean, 2008; Skok & Roberts, 2016). FSS ranges from 0 to 1. A perfect model would yield $FSS = 1$, whereas a model with no predictive skill would yield $FSS = 0$. It is well suited to verifying the accuracy of convection at different spatial scales (Roberts & Lean, 2008). Equation (1) proposed by Lagerquist and Ebert-Uphoff (2022) was used to define the (spatially enhanced) FSS-based loss function in our U-Net architecture:

$$FSS = 1 - \frac{\sum_{g=1}^G [\overline{p_g}(r) - \overline{y_g}(r)]^2}{\sum_{g=1}^G [\overline{p_g}^2(r) + \overline{y_g}^2(r)]}, \quad (1)$$

where $\overline{p_g}(r)$ is the mean predicted value for neighbourhood of radius r , $\overline{y_g}(r)$ is the mean observed value for neighbourhood of radius r , G is the group containing all the neighbourhood windows, and r is neighbourhood radius (spatial scale). The length of the neighbourhood window is defined by r and is equal to the number of pixels in each direction (width and length) centred on the pixel g . We performed several tests to select the most suitable r value and found that a neighbourhood window of 5×5 pixels gave stable results across different evaluation scales. Here, FSS is based on a normalized sum of the squared errors between the mean prediction and mean observation in the neighbourhood of each pixel. Our test results differed from Lagerquist and Ebert-Uphoff (2022) in that their radar-based inputs were able to give similar accuracy values for pixel-wise loss functions compared with spatially enhanced loss functions at 1 hr lead time. However, we observed a notable improvement in accuracy for models trained using spatially enhanced FSS with satellite-observed TIR observations as input, particularly for lead times greater than 2 hr. This is possibly due to the difference in the two studies' target lead time. We are focusing on lead times longer than 1 hr as well and Lagerquist and Ebert-Uphoff (2022) note improvement by using spatially enhanced loss functions as compared to pixel-wise for their 2 hrs predictions. For further details about the structure of the loss function, please refer to Lagerquist and Ebert-Uphoff (2022, fig. 6).

2.1.2 | Experimental set-up

A separate model is trained for different lead times to limit the transfer of errors over successive steps. This approach also gives the subsequent model predictions the chance to self-correct using the latest TIR data. The model outputs an image in which each pixel represents the probability of a convective core being present at that location. We carried out a suite of simulations to optimize the model layers and the selection of input images for training. Separate test models were trained for lead times of 15, 30, 60, 120, 180, 240, and 360 min. For brevity, we only discuss the models trained for lead times 1 hr (60 min), 2 hr (120 min), 4 hr (240 min), and 6 hr (360 min) in this study. Here, we consider lead times of 1 and 2 hr as short and lead times > 2 hr as long. This categorization is to assess the model performance during near versus distant future. Further details about the model development process are described in Appendix A.

2.2 | Study area

The model is trained and then tested for three domains within Africa (which is our main study region) that represent a variety of climates and elevation changes across the African continent (Figure 1b). To maximize convective core frequencies in the dataset, the peak rainy months are selected, which are different for each domain. The first domain is in southern Africa (SA; centred over Zambia) and has a climate that is characterized by three seasons depending on temperature and precipitation. In this study, we look at the peak rainy months of December, January, and February for this domain. The second domain lies in eastern Africa (EA) and is centred around Kenya. It includes Lake Victoria and captures the lake breeze and land interactions near the large water body and the high-elevation regions surrounding the lake. Multiple factors, such as orographic lifting, sea-breeze from the ocean, and high potential evapotranspiration, contribute to convective activity in this region. The peak rainy months are April and May here. The third domain is in western Africa (WA). It focuses on capturing the large mesoscale storms that are typical of the West African monsoon. It also includes the interactions between land and the large coastline on the western and southern edges. The elevation across this whole domain is quite low compared with the other two regions of interest. The three domains in Figure 1b represent varying climates and have different convection drivers and, hence, types of convection. Thus, the model is tested for its ability to nowcast convective cores associated with scattered as well as long-lived convection.

2.3 | Training data

The model training data consist of TIR T_b as input and convective core binary masks as target. NetNCC is trained using input images of 512×512 image width and height. Separate models were trained for each domain based on the peak rainy months in that region. All the models were well-trained using 15 years (2005 to 2019) of data, and testing was done using 4 years' data (2020–2023). The training data were further divided into training (80%) and validation (20%) datasets. For the WA domain, the model was trained and tested using data from July to September. For the SA domain, the training and testing were done using data from December to February. Data for the months of March to May was used for the EA domain. Thus, the models are well trained on 45 months (96 images per day at 15 min temporal frequency) of data and tested in detail using 12 months of data for each lead time. The predictions are probabilistic with no noticeable spread if an ensemble of predictions is made using the same input images. The following sections describe the training and testing datasets in further detail.

2.3.1 | MSG TIR imagery

The spinning enhanced visible and infrared imager (SEVIRI) of the geostationary MSG satellites collects data in 12 channels for multiple applications, such as cloud imaging and tracking, fog detection, and surface and cloud-top temperatures (Aminou, 2002). For this study, we use the TIR effective radiance data in channel 9 ($10.8 \mu\text{m}$) available at the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) website (<http://www.eumetsat.int>). Effective radiances were converted to T_b using conversion tables provided by EUMETSAT. Images are available in near real time at a temporal resolution of 15 min and at an approximate spatial resolution of 3 km in WA.

The raw TIR T_b values were first filtered using a threshold of 0°C to focus on the cloud-covered pixels only. Then the values were normalized using a minimum value that covers the full range of observed T_b . The normalization yields values between 0 and 1 such that very cold clouds had magnitudes closer to 1; see Figure 1a. For consistency, the same normalization threshold was used for all domains. No regridding was applied and the images were displayed on their native MSG grid. Since the input image size was restricted to 512×512 , it meant that a larger area-on-ground was covered for the EA domain using the same number of MSG pixels. Using the native MSG grid helped in training and testing the model's nowcasting skill at a high spatial resolution ($\sim 3 \text{ km}$), meaning

that small-scale convective cores were also captured in the training data. Separate experiments were conducted to train the model at a regular 0.04° grid, and it was noted that the model performance was not impacted by the use of an irregular (MSG) versus a regular grid. Further details about the MSG grid are available in the user data guide (EUMETSAT, 2025).

2.3.2 | Convective cores

The target data for NetNCC comprises binary masks of convective cores. Klein *et al.* (2018) used wavelet scale analysis to identify convective cores within mesoscale convective systems using satellite-based TIR imagery (in our case it is the same input TIR data). The study showed that large mesoscale convective systems consisting of cold clouds are responsible for 91% of all extreme rainfall events in West Africa. Thus, convective cores represent that part of the mesoscale convective system where updraught velocities are high and where deep convection takes place that is usually associated with extreme rainfall (Klein *et al.*, 2018).

Following Klein *et al.* (2018) for convective core detection, cloud-top temperatures below -40°C are identified from MSG images and a wavelet scale decomposition is applied, covering scales from 9 to 130 km via a two-dimensional continuous wavelet transform (Marr & Hildreth, 1980). The wavelet power values are summed over the different scales considered, with convective structures identified where the summed power values exceed a scale-dependent threshold. In this work, the minimum power threshold is defined as $\sum (s)^{0.5}$, where s corresponds to the wavelet scale. Summing over the square root ensures a stronger weighting of small scales within the organized cold-cloud envelope, which are associated with higher probabilities for extreme rain than larger scales (Klein *et al.*, 2018).

For this study, we convert the resulting convective core wavelet transform powers into a binary mask that only provides information about the presence or absence of a core at each pixel. Hereafter, we refer to convective cores as “cores” in this article. NetNCC is trained using the core mask as a target to predict the probability of a core being present at each pixel at the prescribed lead time. There is no explicit difference in the model between convective initiation or propagation of a pre-existing core in the image. The model nowcasts the probability of cores such that larger probability values indicate higher likelihood of a core being present (as predicted by the model).

NetNCC uses a U-Net architecture for two-dimensional images, and thus prediction accuracy is expected to vary across the image based on the frequency

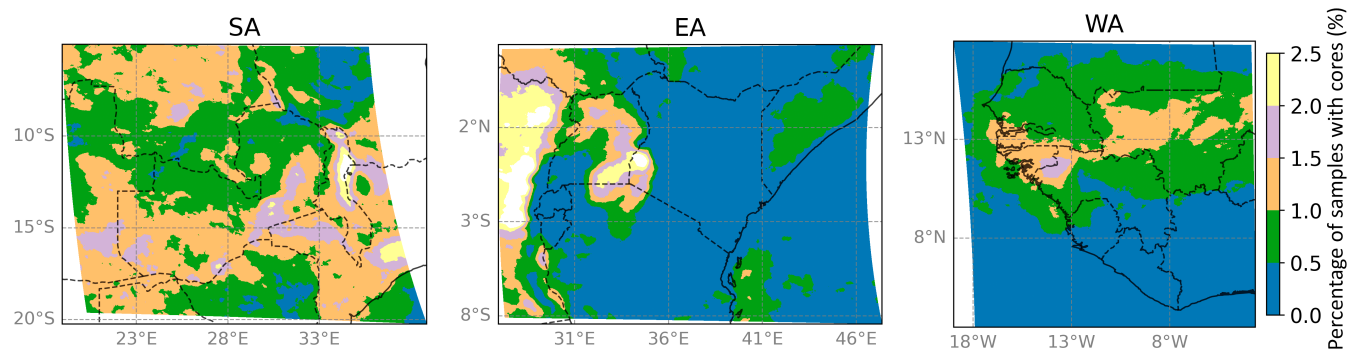


FIGURE 2 Spatial variation in the frequency of cores represented as monthly averaged percentage of pixels with occurrence of cores for each domain. SA: southern Africa; EA: eastern Africa; WA: western Africa.

of core occurrence at each pixel in the training data. Figure 2 shows the monthly averaged percentage of core occurrence at each pixel for the three study areas. Comparing these maps with the terrain elevation maps in Figure 1, it is notable that relatively higher percentages of cores occur over large water bodies (such as Lake Malawi in SA and Lake Victoria in EA) and areas with topography and large elevation gradients. In the WA map, a higher frequency of cores occur near the coast of Guinea-Bissau and Guinea, indicating the interaction between sea breeze and large elevation gradients in that region, and across Mali where long-lived mesoscale convective systems often occur during the rainy season (Zipser *et al.*, 2006). The prediction accuracy of deep-learning-based models is significantly dependent on training data. It is expected that the model would perform relatively well for regions where the training data include a larger number of samples with core occurrences.

3 | RESULTS

In this section, we evaluate the model's nowcasting performance using pixel-wise metrics as well as metrics that account for spatial features at different scales. Seasonally averaged results are presented here that show the model performance across each domain at four lead times for the 12 test months. Each month had more than 2,500 sample images for testing. Thus, the metrics are based on a large dataset and highlight the overall accuracy of NetNCC.

Figure 3 presents an example from the EA domain of NetNCC TIR input images and nowcasts for all lead times in comparison with observed cores. Locations of cooling (increasing normalized values) and warming (decreasing normalized values) TIR T_b correspond well with locations of high core probability magnitudes, such as cores in the western part of the image. However, the model incorrectly predicts the expansion of cores in the top right corner at

lead times 4 and 6 hr. Note that the model does not predict convective core activity in the presence of the extensive cold cloud approaching the Kenyan coast in the south-east of the image, consistent with subsequent observations. The model captures the cores at lead times 1 and 2 hr very well in terms of shape and location. For most test cases, at longer lead times, we noted an increase in the number of pixels with low probability magnitudes with a decrease in the accuracy of predicted core shape as well as location. Similar results were obtained for the other two domains.

3.1 | Evaluation FSS

The evaluation FSS is calculated using Equation (1) to infer the nowcasting accuracy at different spatial scales. In this case, the spatial case is defined by the filter window used to compute the metric. Similar to Lagerquist and Ebert-Uphoff (2022), we do not convert the nowcasts into binary masks for evaluation and instead use the probability values to compute the statistic. FSS ranges between 0 and 1 such that 0(1) indicates no(perfect) forecast skill. Evaluating across scales helps in determining the granularity of the nowcasting accuracy. The TIR input data for NetNCC has high spatial resolution; therefore, nowcasts are provided at a fine spatial resolution that can resolve small cores as well as large convective systems.

Figure 4 shows FSS values for the three domains as the evaluation spatial scale is increased from a filter window of 3×3 pixels ($\sim 9 \text{ km} \times \sim 9 \text{ km}$) to 99×99 pixels ($\sim 300 \text{ km} \times \sim 300 \text{ km}$). We adopt the rationale of Roberts and Lean (2008) for the FSS to be considered skilful above a minimum threshold that accounts for climatology:

$$\text{FSS}_{\text{skilful}} = 0.5 + \frac{\text{Frequency of cores}}{2}. \quad (2)$$

This rationale is commonly used for precipitation nowcasting. However, in our case the frequency of cores is quite

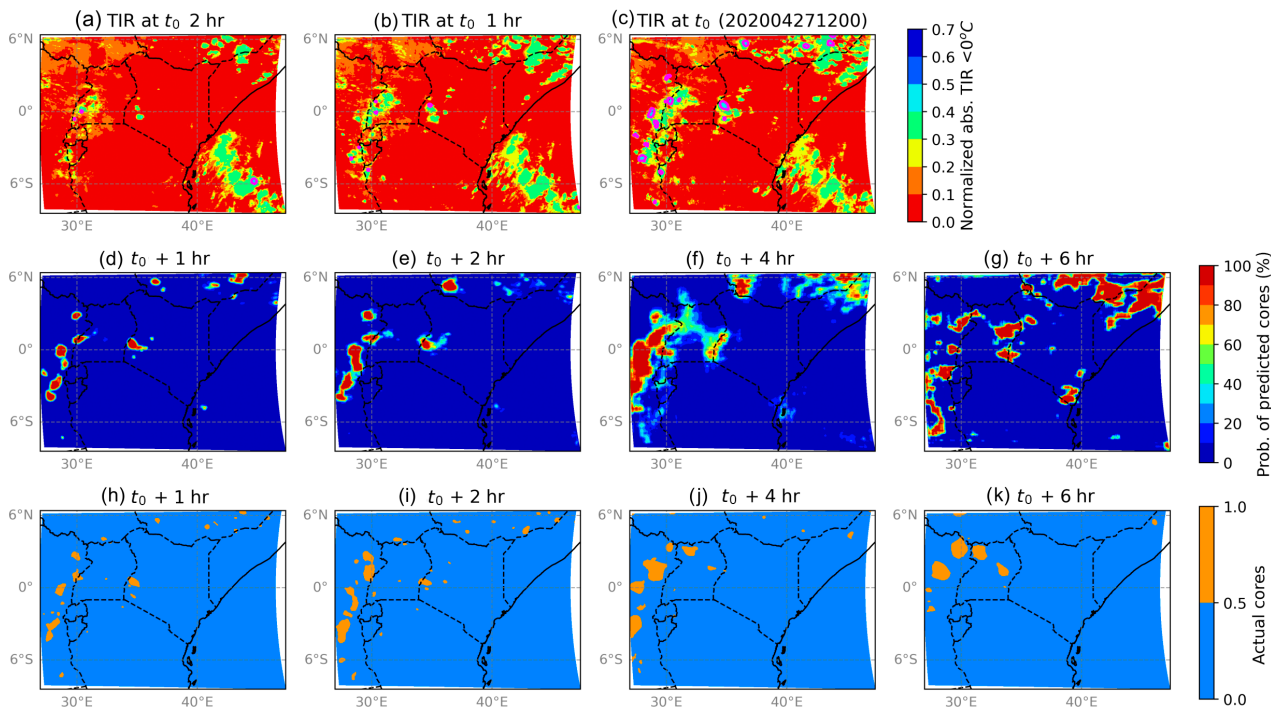


FIGURE 3 Network for Nowcasting Convective Cores (NetNCC) nowcasts initialized on April 27, 2020, at 1200 UTC (1500 h local time) for the eastern Africa domain for different lead times. (a)–(c) The thermal infrared (TIR) input images. Magenta contours on TIR images indicate cores at each time step as detected by the core algorithm but not provided as NetNCC input. (d)–(g) The probabilities of cores predicted by NetNCC based on the TIR imagery. (h)–(k) The true core mask at each lead time.

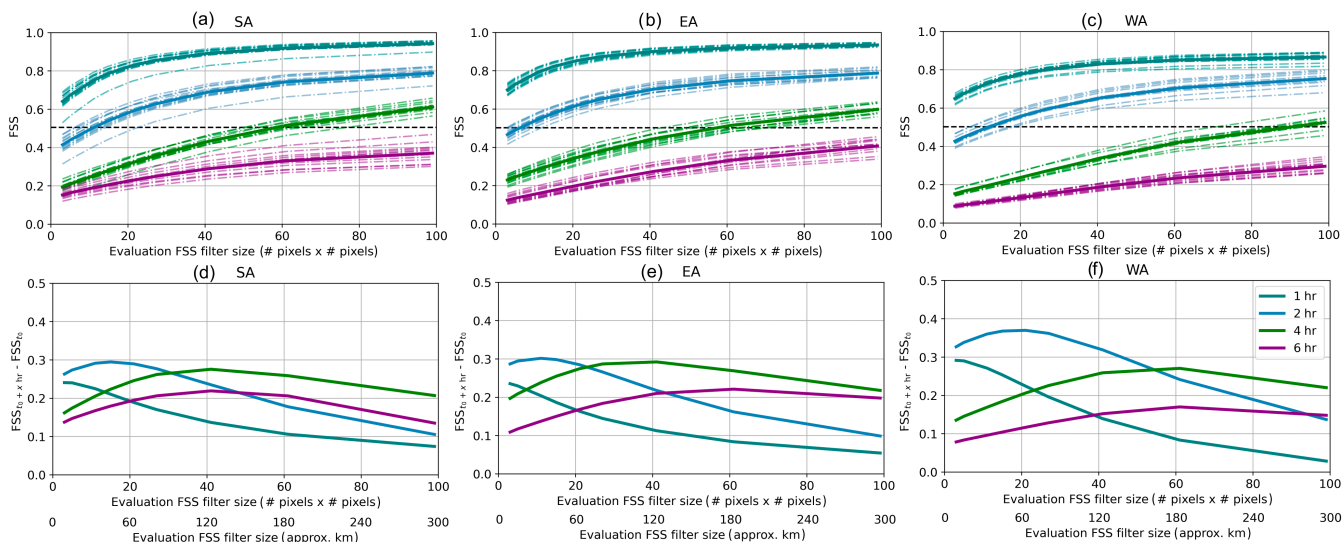


FIGURE 4 (a)–(c) Fractions skill score (FSS) for three domains at increasing spatial scales for different lead times. The black dashed line represents minimum FSS threshold for skilful prediction: $0.5 + (\text{Frequency of cores}/2)$. Each individual dashed line represents FSS computed for predictions from one test month (total 12 months), and the solid line is for mean of all months. (d)–(f) The differences between FSS for nowcasts at lead time ($t_0 + x$ hr) and persistence (t_0). Skilful predictions are obtained at fine spatial resolution for short lead times (≤ 3 hr) and at coarser resolution for longer lead times (> 3 hr). SA: southern Africa; EA: eastern Africa; WA: western Africa.

low and does not impact the minimum threshold much. We find that FSS values increase across all domains as the evaluation spatial scale is increased for all lead times

(Figure 4a–c). As expected, FSS decreases with lead time at all spatial scales. For shorter lead times (≤ 3 hr), the plot slope is steep for smaller scales and then becomes flat

as the spatial scale becomes coarser, indicating no gain in skill by further coarsening of nowcasts. However, for longer lead times (>3 hr) FSS values increase at a similar rate for finer and coarse spatial scales. The range (variability) of accuracy for different test months is represented by the spread of dashed lines, which is smaller for shorter lead times and larger for longer lead times. This behaviour is noted for SA and EA domains, whereas it is less clear for WA, where the dominance of long-lived convection and large interannual variation in monsoon band positioning could make differences in lead time spread less pronounced.

Skilful predictions are obtained for lead time of 1 hr at all scales for all domains. The model is robust in terms of 1 hr prediction at quite a high resolution (~ 9 km), finer than most contemporary nowcasting models. For lead time of 2 hr, the skilful prediction threshold is crossed at 5×5 pixel neighbourhood (~ 15 km) for EA and 10×10 (~ 30 km) pixel window for SA and WA. The skilful prediction threshold for lead time of 4 hr is crossed at a much coarser resolution; that is, the 60×60 pixel window (~ 180 km) for SA and EA and the 90×90 pixel window (~ 270 km) for WA. The nowcasts at lead time of 6 hr do not cross the skilful accuracy threshold even at the large scales analysed, though nowcasters may find useful large-scale information at this lead time, particularly in EA, where $FSS > 0.4$. In general, the best FSS values are noted for the EA domain at lead times 1 and 2 hrs; for a lead time of 4 hr the EA and SA domains have quite similar performance, and the SA domain has slightly better accuracy than the other regions at lead time of 6 hr at smaller spatial scales. The WA domain exhibits a lower accuracy than the other two domains at all lead times. These scores are affected by the type and frequency of convective cores that occur within the domains. Thus, it seems that NetNCC is better able to predict the generally short-lived convective cores that develop in SA and EA in the presence of strong topographic forcing.

Next, we consider NetNCC nowcast skill relative to assumed persistence of the atmospheric field (Figure 4d–f). In meteorology, persistence refers to the surface weather or atmospheric circulation remaining in a quasi-stationary state or recurring to a specific state for a long period of time (Tuel & Martius, 2023). Here, we use the first definition and define it as the core retaining the same shape and size and remaining at the same location at the prescribed lead time as time t_0 . Thus, FSS values for persistence are calculated using core masks at time t_0 (FSS_{t_0}). Figure 4d–f shows the resulting difference between FSS for nowcasts at respective lead times ($t_0 + x$ hr) and persistence (t_0); that is, assuming that convective activity stays constant over time. These plots highlight the additional information and added utility

the nowcasts provide compared with using current cores for convection prediction, even for short lead times. Positive differences are seen for all lead times for all three domains; that is, the nowcasts are better than persistence for all domains. The largest differences are observed for the WA domain for shorter lead times (<3 hr), particularly for lead time of 2 hr, with similar differences for SA and EA at all spatial scales.

Large convective systems develop and propagate during the rainy season in WA. The model performance suggests that it is quite capable of nowcasting the propagation of these convective systems in WA for shorter lead times compared with persistence. At lead time of 4 hr, the highest differences are noted for EA for smaller spatial scales; that is, the model is nowcasting the smaller core features better than persistence. Low positive values are seen for all domains for lead time of 6 hr, with relatively higher differences (0.2) for smaller spatial scales in SA and for larger spatial scales for EA. Across all domains, larger differences between nowcast and persistence accuracy occur at finer spatial scales for shorter lead times and coarser spatial scales for longer lead times. In other words, the nowcasts are better than persistence at resolving spatially fine convective features at short lead times and spatially coarse convective cores at longer lead times. Overall, this shows that NetNCC is able to produce skilful patterns of convective activity out to at least 4 hr, where predicted changes in convection extent and location outperform persistence assumptions already at 1 hr.

3.2 | Diurnal cycle prediction accuracy

In Figure 4, we looked at pooled nowcasts made at all times of day. However, given the diurnal cycle in radiative fluxes that produces a pronounced afternoon frequency peak of convection over land, it is expected that the nowcast accuracy would vary throughout the day. Figure 5 shows the change in nowcasting accuracy throughout the day for different lead times as measured by FSS, computed for nowcasts made at each hourly interval. Chosen FSS filter radii were selected based on the slope of the FSS plots in Figure 4 so that a distinction can be made between the nowcasting accuracy across a relatively smaller versus a larger spatial scale. Figure 5e shows the diurnal cycle in the frequency of cores for each domain as represented by the percentage of total sample pixels identified as convective core. It can be seen that frequency of convective cores peaks between 1100 h and 1300 h local time in SA and EA, in contrast to a peak around 1900 h in the WA domain, consistent with a strong relative contribution of long-lived mesoscale convective systems to total rainfall in the latter region.

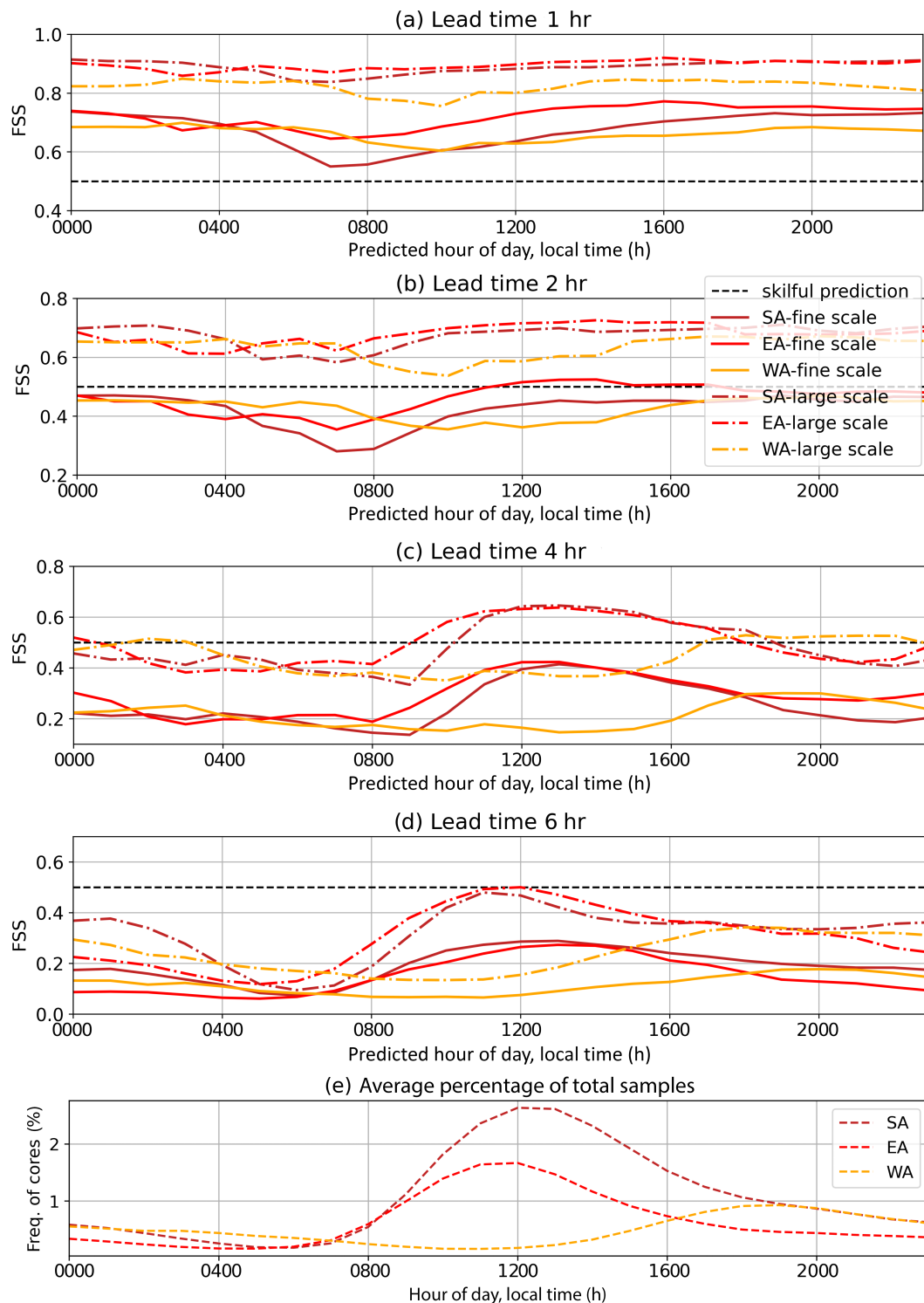


FIGURE 5 (a)–(d) Fractions skill score (FSS) computed for nowcasts at each hour of day at a small versus large spatial scale at different lead times. Solid lines show a fine spatial scale (5×5 for lead times 1 and 2 hr and 21×21 for lead times 4 and 6 hr) and dash-dot lines show a large spatial scale (41×41 for lead times 1 and 2 hr and 81×81 for lead times 4 and 6 hr) for the three domains. (e) The frequency of cores with respect to the local time of day. SA: southern Africa; EA: eastern Africa; WA: western Africa.

FSS values are relatively higher for daytime nowcasts made for 0800 h to 1700 h local time for all lead times for the SA and EA domains. This matches well with the higher

frequency of cores during those hours (Figure 5e). There is a dip in FSS values for the SA and EA domains in the early morning hours for all lead times which begins to increase

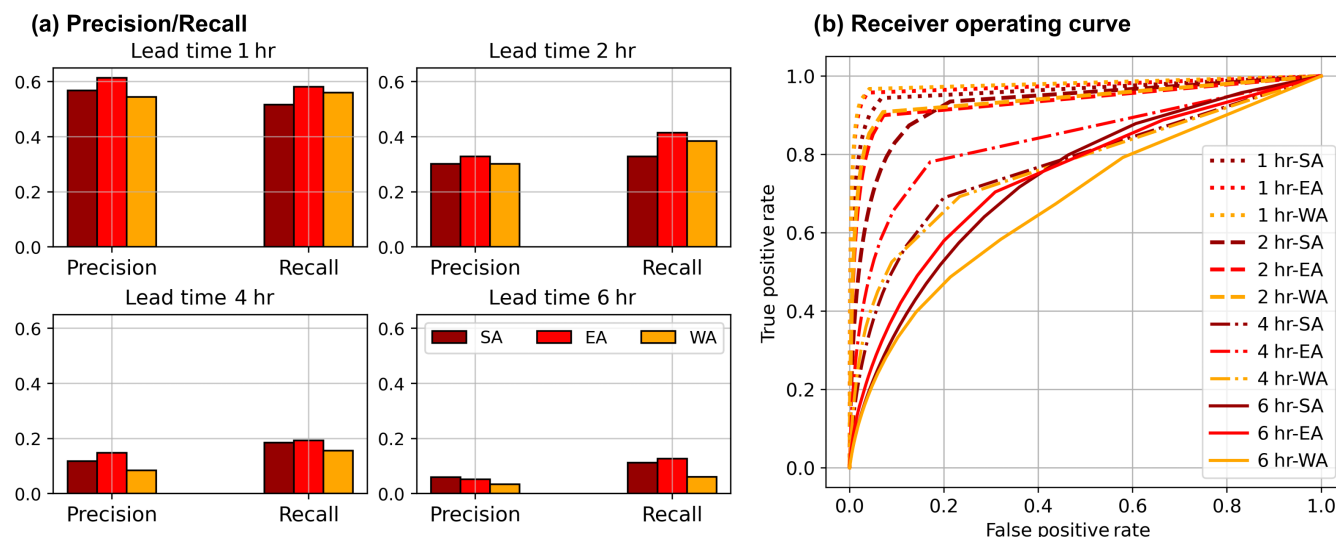


FIGURE 6 (a) Pixel-wise accuracy with respect to lead time for each study area, where Precision = True positives/(True positives + False positives) and Recall = True positives/(True positives + False negatives). (b) Receiver operating characteristic curves show the true (core correctly predicted) versus false (core incorrectly predicted) positive prediction rate for all domains at each nowcasting lead time. The curves highlight the change in true versus false prediction of cores as the probability threshold used for binary core masks is changed. SA: southern Africa; EA: eastern Africa; WA: western Africa.

after 0800 h local time. This trend is generally noted at smaller as well as larger spatial scales. However, this trend is not observed for the WA domain, where the FSS values are slightly lower from 0800 h to at least ~1200 h local time for all lead times. In WA (yellow lines in Figure 5), the nowcast accuracy is relatively higher during the night-time compared with the daytime (increases around 1500 h local time for all lead times). This trend is similar to the frequency of core occurrence in WA, where a slightly higher percentage is noted after 1500 h local time (yellow line in Figure 5e). The WA nowcast accuracy has less variability throughout the day compared with EA and SA, with lower FSS values particularly from 1000 h to 1700 h local time, matching the frequency of core plot.

For lead times of 1 and 2 hr, the FSS plots for EA and SA show an increase and decrease in accuracy at similar times of day for both the small and large spatial scales of evaluation. The only difference is in the magnitude of decrease or increase in FSS values, which is higher at smaller spatial scales, thus exhibiting relatively higher variability in nowcasting accuracy at smaller spatial scales. The spatial scales represent the adequacy of the nowcasts in capturing fine- and large-scale convective features. Thus, it can be deduced that the model nowcasts are better able to capture fine-scale convective features during the daytime (compared with the night-time) in SA and EA, when most of the convection events occur. There is a marked improvement in the FSS values computed during daytime at lead times 4 and 6 hr for SA and EA domains at both spatial scales. For a lead time of 4 hr, the nowcasts cross the skilful prediction

threshold at the larger spatial scale, whereas for a lead time of 6 hr the FSS values almost reach the skilful prediction threshold around noon local time. During night-time, the FSS values are relatively lower (compared with daytime) and show less variability for all domains.

3.3 | Pixel-wise evaluation

So far, we have discussed nowcasting accuracy for different filter windows (i.e., spatial scales). In this section, we look at the pixel-wise (MSG native grid) predictive accuracy of NetNCC. Figure 6a presents the mean precision and recall values for each study area. Precision reflects the accuracy of the model, whereas recall represents the ability of the model to predict a relevant class. In our case, precision is the proportion of predicted core pixels that are actually cores, and recall is the proportion of core pixels correctly predicted as core. An ideal model would have high precision and high recall, thereby correctly identifying core occurrence. As expected, the precision and recall decrease with lead time for all domains. For longer lead times, the precision is small as the number of false positives is high. A relatively higher number of pixels are classified as core at longer lead times by the model (see Figure 3d–g), which leads to an increase in false positives. For a lead time of 1 hr, the mean precision is higher than recall, whereas for lead times >1hr the recall is slightly higher than precision. On average, precision and recall values for EA are relatively higher, followed by SA in terms of precision. Note

that most of the predictions made by the model are true negatives, meaning that the model is able to correctly predict the absence of cores the majority of the time. On a pixel-wise basis, the model's accuracy is more similar for all domains compared with the previous neighbourhood FSS analysis.

Figure 6b shows the receiver operating characteristic (ROC) curves for all lead times. ROC curves are commonly used in meteorological forecasting to look at true- versus false-positive prediction rates for a binary classifier at different thresholds (probability values in this case). The goal is to have a model that has high true-positive rates for low false-positive rates such that the area under the curve is large. At 1–2 hr lead times, the EA and WA domains have higher true-positive rates for the same false-positive rate than the SA domain, with WA showing slightly higher rates. However, for longer lead times (>3 hr) the EA nowcasts (yellow lines) show better performance than the other two regions. Different from the results using a neighbourhood metric (FSS) sensitive to spatial patterns, the ROC curves suggest that the nowcasts for WA are relatively a bit better than the SA domain at the pixel-wise level at short lead times, when potential displacement of convection is still small. However, the skill for WA decreases markedly at longer lead times, aligning with the FSS results. Though one might expect the longer-lived convection in WA to enhance predictability at extended lead times, the lower skill seen in both ROC curves and FSS indicates that accurately displacing organized but often isolated systems is more challenging for NetNCC than nowcasting more frequent, scattered convection tied to persistent landscape features in EA and SA.

3.4 | Spatial prediction accuracy

Regional circulation patterns and topography often affect the development of convective cores at fine spatial scales. Additionally, for an image-based deep-learning model, it is expected that the predictability of cores will vary across different parts of the image depending on the spatial features inherent in the training data. Here, we use the Brier skill score (BSS) to evaluate the model's spatial accuracy. BSS has extensive usage in verifying probabilistic predictions in meteorology (Brier, 1950; Weigel *et al.*, 2007). We use a modified version of BSS to account for persistence instead of climatology to set a more stringent baseline for the nowcasts. Equation (3) is used to compute Brier Score and BSS:

$$\text{Brier score}_{\text{pred}} = \sum_{n=1}^N [y_{\text{pred}_n} - o_n]^2, \quad (3)$$

$$\text{Brier score}_{t_0} = \sum_{n=1}^N [y_{t_0_n} - o_n]^2, \quad (4)$$

$$\text{BSS} = 1 - \frac{\text{Brier score}_{\text{pred}}}{\text{Brier score}_{t_0}}, \quad (5)$$

where y_{pred} is probability of predicted core, y_{t_0} is persistence (core at time t_0), o is the observed core (0 or 1), n is the sample (image) number, and N is the total number of samples (images for evaluation). BSS values range from 0 to 1. We are using BSS to measure the relative skill of the forecast compared with persistence. Therefore, BSS = 1 (0) means that the nowcast has perfect (no) skill compared with persistence, whereas a negative BSS value means the nowcast is less accurate than persistence. The goal is that the Brier score should be near 0, whereas the BSS should be near 1 for perfect predictions.

Figure 7 shows BSS spatial maps for the three domains at four lead times. The maps show BSS values for nowcasts smoothed to a spatial scale of ~45 km (filter window equal to 15 × 15 pixels). The general accuracy as depicted by BSS values is high (> 0.3) for lead times up to 4 hr and decreases relative to lead time for all regions. As expected, BSS values are highest for areas where the average frequency of cores is large (comparing Figures 2 and 7), partly by giving the model more opportunity to learn from the training data and to correctly calibrate probability values.

Overall, the maps indicate that the model is able to learn about the impacts of local topography and water bodies on convection. For example, there are high values of BSS over and near large water bodies (Lake Malawi in SA and Lake Victoria in EA). Similarly, higher BSS values are noted for locations where sharp elevation gradients are present in all three domains (such as the western part of EA domain near the DRC border and over Guinea in WA). An interesting location where the model performs well for all lead times in SA is the valley between longitudes 29°E and 31°E. It lies between two sharp elevation gradients which facilitate orographic lifting leading to convective activity. The sensitivity of nowcasting accuracy to location is clearer at lead time of 6 hr. It is worth noting that even for short lead times of 1 and 2 hr the model's nowcasting performance is considerably higher than persistence (BSS > 0.5). It indicates the added value of the nowcasts at such short time periods given the rapid evolution of convective structures.

There are a few locations where the model yields negative BSS values meaning that it performs worse than persistence, such as the lower part of the WA domain. There are very few convective cores (as defined in our study) that occur in this region and thus it is expected that the model's nowcasting accuracy would be limited in that

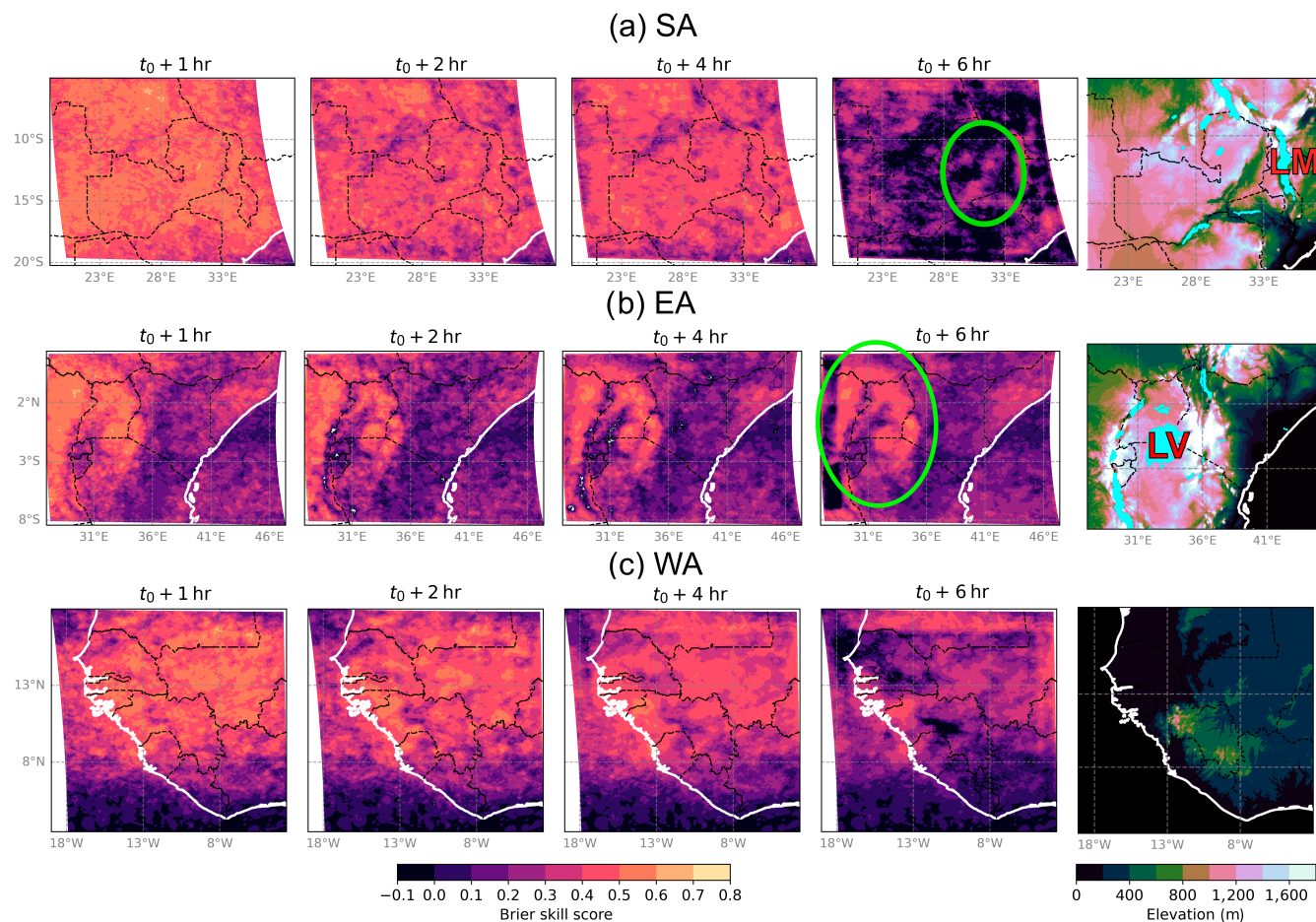


FIGURE 7 Spatial maps showing the accuracy of the nowcasts compared with persistence using Brier skill score (BSS) for different lead times ($t_0 + \text{lead time}$). The last column shows elevation maps. The spatial variability in BSS values is similar to the variability in core occurrence frequency (Figure 2). Large BSS values indicate the locations of high nowcasting accuracy for the model (such as encircled areas that correspond with elevation gradients). SA: southern Africa; EA: eastern Africa; WA: western Africa; LM: Lake Malawi; LV: Lake Victoria. White lines show coastline, and lakes are shown in cyan.

region. The spatial maps also show some model deficiencies along the edges. Model performance is limited due to common image segmentation errors such as edge effects and kernel issues and missing information on existing convection outside the domain which become more important for longer lead times (e.g., in WA at lead time of 6 hr). However, these errors are small in magnitude and can be removed using a first-order approach such as spatial 2-D convolution using a small filter window.

Another important point highlighted by the spatial maps of BSS is that while the general FSS magnitudes were comparably low for WA (Figure 4), the BSS values are high even for lead time of 4 hr across the majority of the WA domain (Figure 7). Considering the long-lived nature of convective storms in WA during July to September, persistence could yield adequate accuracy at least for short lead times. However, we note in Figure 8 that Brier score values for nowcasts are much lower as compared to persistence. Brier score values for nowcasts are larger

in magnitude for locations with high inherent frequency of core occurrence. However, the Brier score values for persistence are even higher, thus resulting in positive BSS values. In addition to the general BSS maps included in this section, we also explored spatial accuracy of nowcasts for different times of day. The results are presented in Appendix B and are provided to help operational forecasters better understand the performance of the nowcasts in different domains throughout the diurnal cycle.

3.5 | Comparison with existing nowcasting methods

Anderson *et al.* (2023) predict the probability of convective cores at lead times of 0 to 6 hr using a historical climatology, conditioned on the location of real-time cores. The method was developed as part of the Nowcasting Flood Impacts of Convective storms in the Sahel (NFLICS).

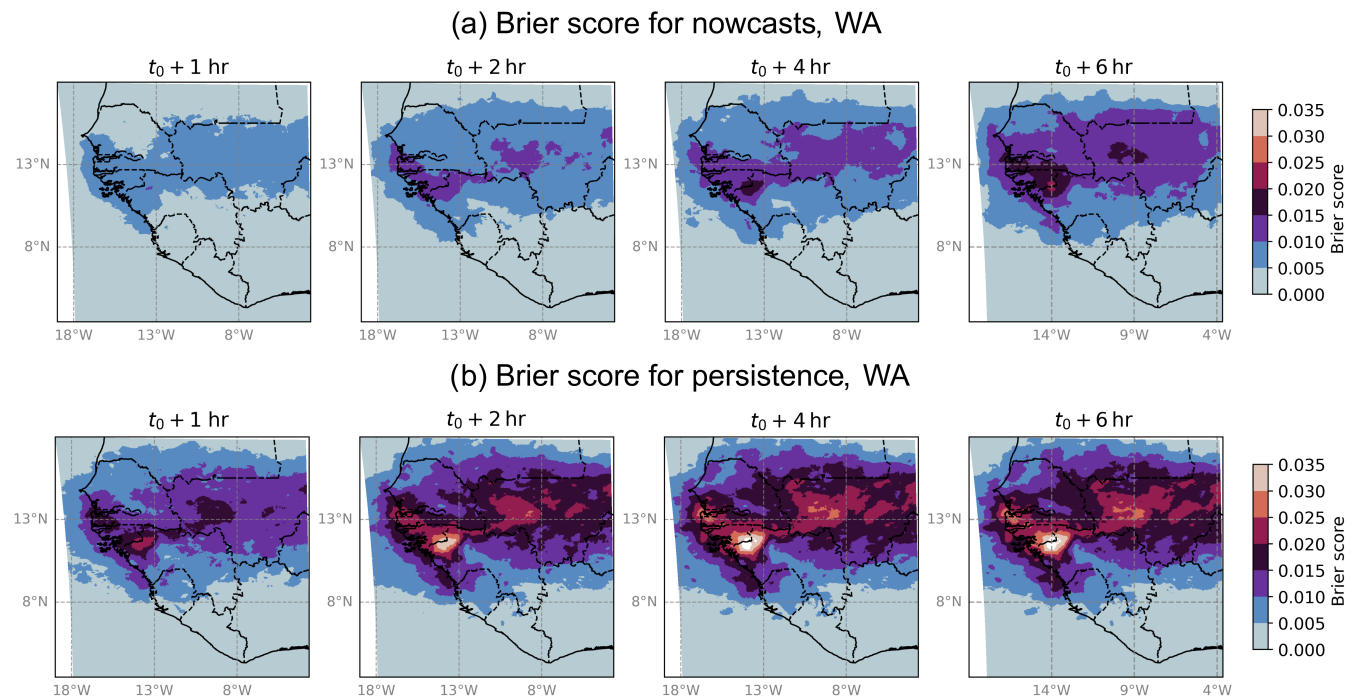


FIGURE 8 (a) Spatial maps showing Brier score computed for nowcasts at each location for the western Africa (WA) domain for different lead times. (b) As (a) except for persistence; that is, cores at time t_0 .

NFLICS predicts the probability of core occurrence at different spatial scales (neighbourhoods) surrounding each grid point. The method was shown to be skilful, particularly at longer lead times. The primary dataset used for NFLICS and NetNCC is quite similar; that is, TIR T_b . Here, we compared the nowcasts produced by NFLICS with NetNCC for July, August, and September of 2020–2022. The NFLICS nowcasts were provided at a 5 km regular grid and are regridded on the MSG-based NetNCC grid (~ 3 km) for comparison.

Figure 9 shows the Brier score computed for NetNCC and NFLICS nowcasts at lead times 1, 2, 4, and 6 hr for the WA domain. The scores were computed at a spatial scale of 45 km to account for displacement errors and limit the double penalty effect. NetNCC outperforms NFLICS at all lead times considered, with much lower Brier scores. Spatially, NetNCC nowcasts are constrained by local convection and less tied to the climatological frequency of core occurrence. NetNCC nowcasts exhibit better spatial sharpness, with higher probability magnitudes predicted over localized areas driven by cold cloud temperatures at the nowcast initiation time. In comparison, NFLICS predicts lower probability magnitudes over larger areas surrounding the cold cloud temperature pixels. Figure 9a–d shows the general increase in Brier score computed for NetNCC as lead time increases, particularly for regions with higher frequency of cores (Figure 2c). The values are much lower for a lead time of 2 hr relative to 6 hr, as is reflected by

the FSS values in Figure 4. The relative differences in Brier scores for different lead times in NFLICS are more muted. The spatial pattern in Brier score values for both models is similar to the spatial variability in frequency of core occurrence (Figure 2c). This is more noticeable for NFLICS, where we note higher Brier score magnitudes for areas of different percentages of core occurrence in the dataset. Historical climatology is an important part of NFLICS nowcasts, and the similarity in spatial patterns of accuracy and frequency of core occurrence is expected.

An indirect comparison with the methodology used by Burton *et al.* (2022) shows skill at similar spatial scales. Burton *et al.* (2022) used pySTEPS (Pulkkinen *et al.*, 2019) to nowcast satellite-derived convective rain rate in West Africa and showed that mean FSS values greater than 0.5 showed skill in convective rain rate nowcasts to around 1.5 hr at 8 km and to around 2.5 hr at 64 km scale. They used an optical flow method to propagate convective rain rates derived from several MSG SEVIRI channels in the visible to infrared wavelengths (Lahuerta, 2021). The spatial scales at which they obtained skilful nowcasts are similar to NetNCC for 1 hr (< 9 km) and 2 hr (~ 30 km) in WA. Burton *et al.* (2022) reported a diurnal variation in skill with the lowest performance in nowcasts that are initialized at 1500 UTC. In contrast, we note better performance as represented by FSS values for nowcasts initialized in the afternoon (Figure 5), although the Brier scores are higher in the evening in WA.

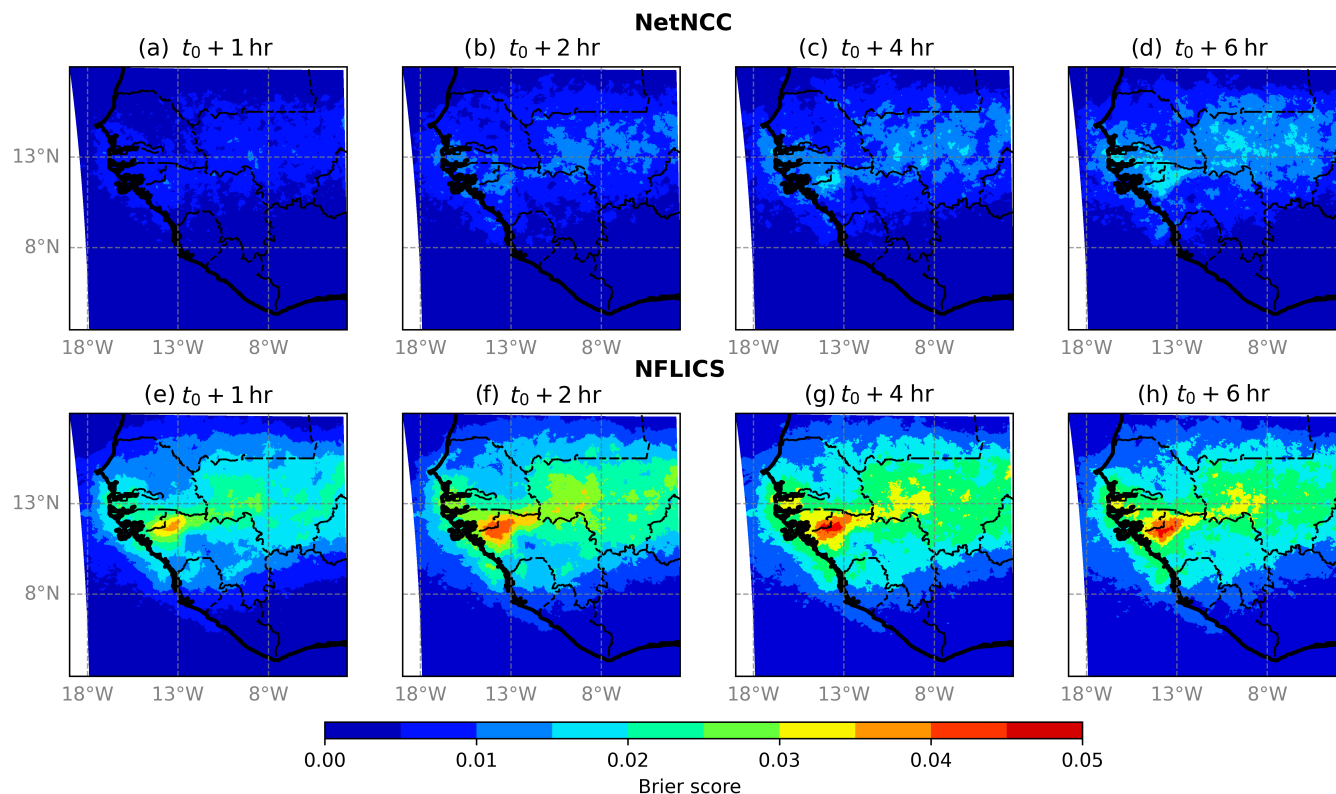


FIGURE 9 Brier score computed for (a)–(d) Network for Nowcasting Convective Cores (NetNCC) and (e)–(h) Nowcasting Flood Impacts of Convective storms in the Sahel (NFLICS) in western Africa at ~ 45 km spatial scale. Predictions for July, August, and September from 2020 to 2022 were used to compute the monthly averaged accuracy.

3.6 | Real-time evaluation during testbed

Along with the statistical metrics computed using historical data, NetNCC was also tested in real time during a 2-week “Forecasting Testbed” held in Lusaka, Zambia in January–February 2025. This event was organized by the Early Warning for Southern Africa project under the UK Weather and Climate Information Services (WISER) (UK Met Office, 2025b) programme and provided an opportunity to pilot nowcasting tools in real time. Further details about the testbed are available on the WISER testbed website (UK Met Office, 2025a). NetNCC nowcasts are available in near real time on the UK Centre for Ecology and Hydrology’s nowcasting portal (<https://africa-hydrology.ceh.ac.uk/nowcasting/>). These nowcasts were evaluated at subhourly, hourly, and daily temporal scales. The nowcasts were able to capture medium- to large-scale convective cores very well, particularly at short lead times. At longer lead times, the nowcasts usually predicted non-zero probabilities in the general area where cores eventually occurred. The participating forecasters used NetNCC in addition to a suite of other satellite-based products and found it more useful for the 2 hr impact-based nowcast. The nowcasting performance was especially robust in regions of high precipitation where the storm impact was

significant, highlighting the considerable potential that NetNCC has for impact-based forecasting.

We include one case of using NetNCC to track hourly storm development during the testbed in Figure 10. In this case, a convective system developed as the sea breeze from the ocean moved inland over Mozambique. The first row shows cloud-top temperatures (T_b). At 1100 h local time (UTC + 0200), cold clouds were already observed at the coastline near latitude -19° S and longitude 37° E. The storm matured and propagated over the next couple of hours such that a large, cold cloud system was noted around 1300 h local time. Comparing the nowcasted versus actual cores, it was seen that core development and propagation was well predicted by NetNCC at the 1 hr lead time. The nowcast probability magnitudes increased as the storm matured from 1400 h local time onwards. It is interesting to note that the two cores at 1600 h local time were well predicted 1 hr prior as the T_b at 1600 h local time shows two distinct cold regions. The evolution of cores was well captured, showing that the model implicitly learned the physical processes that contribute to the expansion and inland propagation of the sea-breeze-driven convective system. NetNCC is developed to help forecasters in operational nowcasting, and its utility in tracking convective storms in real-time was well tested during the testbed.

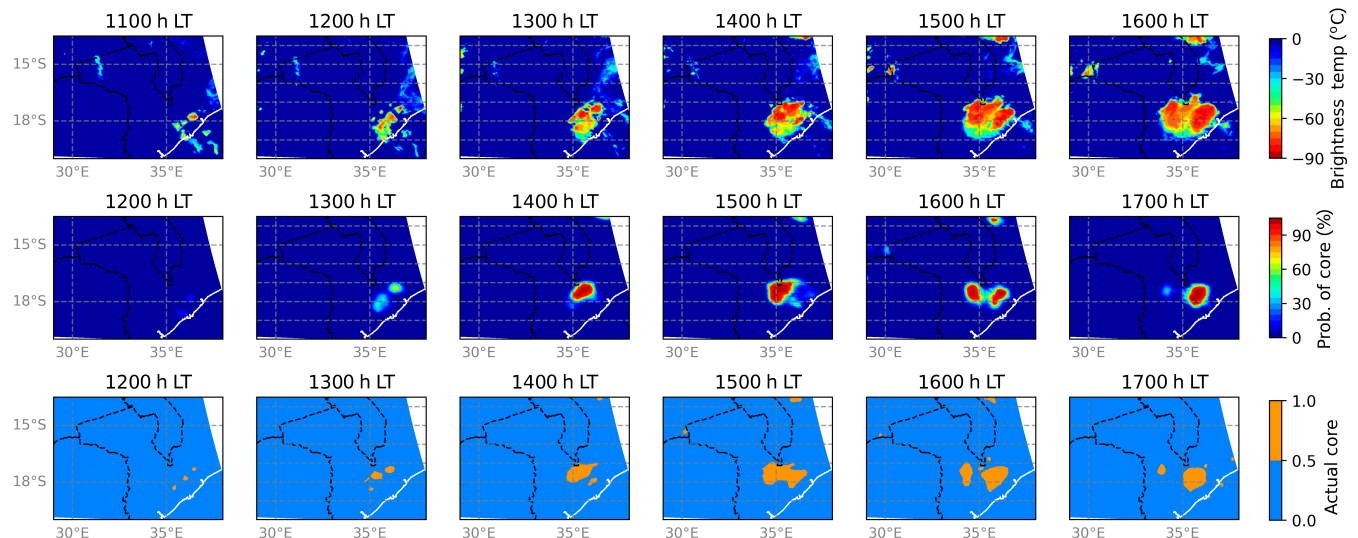


FIGURE 10 Network for Nowcasting Convective Cores (NetNCC nowcasts for lead time of 1 hr for a convective system that developed on the coast of Mozambique on February 2, 2025. The white line shows the coastline. Map titles show the local time (LT; UTC + 0200) in Mozambique. Thermal infrared (TIR) brightness temperatures in the first row highlight the location of cold clouds, and the second row shows the probability of cores nowcasted using the TIR brightness temperatures. Comparison of core probabilities with the actual cores observed at the time shows that NetNCC captured the evolution of the convective system as it moved northwest due to the interaction between sea breeze and land surface.

4 | DISCUSSION

NetNCC was developed with the motivation to use as few as possible near-real-time observations to adequately nowcast convective cores at short time ranges at a high spatial resolution using a simple model architecture. These goals were mostly achieved as detailed by the evaluation in Section 3. A simple and easy-to-train model with adequate accuracy was developed. The primary source of skill in NetNCC is the high-resolution TIR T_b input data and the custom loss function. The loss function we use for training is well suited for the type of spatial features the model attempts to predict. The model implicitly learns about atmospheric wind direction as the TIR T_b values move across the image and become colder or warmer within the successive time step images over the last couple of hours. The model further learns about the underlying climatology of convective systems from the 15 years of training dataset. Thus, such a simple model is able to provide skilful nowcasts up to 4 hr ahead. The learned climatology becomes more important at a lead time of 6 hr, as apparent from the accuracy maps in Figure 7.

The results also highlight some limitations in the described model architecture and nowcasting accuracy. First, NetNCC is dependent on the availability of near-real-time TIR radiances. In the case that TIR images become unavailable or there is a disruption in EUMETSAT data transfer, the model would have no operational nowcasting capability. As with most deep-learning models,

training NetNCC required considerable computational memory. However, once trained, the model predictions are made in seconds. In addition, the model does not achieve skilful predictions for a lead time of 6 hr even at large spatial scales. Some cases show that irrationally high core probabilities that span large areas are predicted along the western edge of EA domain for a lead time of 6 hr. Similarly, for WA, edge-effect-based errors were observed at a lead time of 6 hr along the top edge of the domain where the training data have very few cores from which to learn. Surprisingly, it was also seen that using only two months (for 15 years) of data to train the lead time of the 6-hr model yielded better results than using three months (for 15 years) of data to train the model, which showed local biases along the edges. We tried to remove the bias using (a) a larger filter window, (b) a pixel-wise loss function, and (c) a larger filter window in conjunction with a pixel-wise loss function for a model pretrained with a spatially enhanced FSS loss function. However, using a solely pixel-wise or pixel-wise combined with spatially enhanced loss function for a lead time of 6 hr model reduces the model's ability to predict almost any cores. Further work is needed to improve the model's nowcasting accuracy at a lead time of 6 hr, which could potentially involve using additional channels with more persistent predictive power as input, such as land surface temperature (Taylor *et al.*, 2022) and wind shear (Maybee *et al.*, 2024).

NetNCC is trained for peak rainy seasons and has lower nowcasting accuracy for the dry months (results not

included). The low statistical accuracy during dry months is due to both the lack of convective core occurrence in the test data for dry months and the model's limited extrapolation abilities. The model learns to emulate atmospheric conditions during the rainy season and, in general, gives similar results across domains in terms of convective core development. One potential way forward to improve year-round nowcasting accuracy would be to train the model using data from all the months, thereby yielding better nowcasting accuracy during the dry months. However, it is expected that using data from all the months would lower nowcasting accuracy during the peak rainy season.

NetNCC has high reproducibility courtesy of its simple model framework. It can be implemented for any domain as long as TIR radiances for that domain are available at a suitable spatial and hourly temporal resolution. In the interest of space, we did not include nowcasts for lead times less than 1 hr. The model is able to predict convective cores at 15 and 30 min with high accuracy (FSS > 0.7) at a very fine spatial resolution (~3 km). The model was tested with different inputs to understand the physical interpretability of the nowcasts (results not included). Nowcasts produced by the original model set-up (three successive TIR T_b images as input) were compared with a test case in which the first two input channels were replaced with the TIR T_b image at t_0 . Thus, the model was provided information that showed no temporal change in the propagation of cold T_b clusters or the cooling or warming of cold pixels. The nowcasts produced by the two models were similar in predicting the location and propagation of large cores due to very cold clouds. However, the test case nowcasts predicted low or zero probabilities for smaller-scale cores. The test results indicated that the prediction of large core propagation was primarily driven by the TIR T_b image at t_0 . The prediction of locally initiating cores was more sensitive to all the successive TIR T_b images that contained information about the temporal and spatial features in TIR T_b values.

We used MSG SEVIRI data in this study owing to its extensive temporal coverage. Meteosat Third Generation (MTG) data became available in 2025. Thus, we do not have enough data for retraining the models for MTG, but we do have enough observations for testing purposes. The model trained using MSG data was tested by using MTG-derived TIR T_b values in a similar wavelength (10.5 μm vs. 10.8 μm). For a direct comparison, we only used MSG and MTG files available at the same time (every 30 min). Both the datasets were regridded to a regular 0.05° grid for consistency. The model was trained using MSG data from July–September 2005 to 2019 and tested separately using MSG and MTG data for July 2025. The results showed good transferability for MTG such that similar nowcasts were made by both datasets driven by

very similar input TIR T_b values (figures not included as the results are visually very similar). MTG-based nowcasts tend to have larger spatial buffers around the locations of true cores with slightly higher probability values predicted. Differences in FSS and Brier score values computed for the two sets of nowcasts are minimal. Considering that MTG data are provided at a finer grid (~2 km) than MSG (~3 km), it is expected that MTG will have higher amplitudes capturing small-scale features such as colder cloud temperatures. Any difference (however small) in the input TIR T_b values tends to impact the nowcast probability at 4 and 6 hr lead time more. Based on the results, it seems that fine-tuning MSG-trained model using MTG data should help eliminate any differences in accuracy.

We note higher variability in nowcasting accuracy across the domain at longer lead times. Moving forward, the model will be tested by adding land surface information at different scales in an effort to improve the nowcasting accuracy across the domain, particularly during the afternoon and evening hours when most of the cores occur Taylor *et al.* (2022). There is a considerable class imbalance within the target data due to the nature of convection occurrence. One possible avenue of future research is the modification of the loss function to explicitly address this class imbalance. It is important to customize the loss function in a way that retains the spatial structure of cores while giving more importance to the instances of core occurrence which are relatively scarce. The current framework only uses brightness temperatures from the 10.8 μm MSG channel since the convective core target data is also extracted from that channel. It is possible that relevant complementary physical information may be obtained from other MSG channels as well, particularly from the visible bands during daytime. A multi-channel approach may improve the domain-wide accuracy by providing complementary information. In that case, the channel selection would need to consider the ability of radiation at a particular wavelength to characterize convective core development or propagation. In addition, modifying the model target data to focus on convection events that result in extreme precipitation would advance impact-based forecasting capabilities of the model. For at-risk regions, where early warning systems for extreme weather do not exist at present, a satellite observation-based model like NetNCC could provide considerable socioeconomic benefit.

5 | CONCLUSION

Nowcasting convective cores is important for extreme weather prediction and early warnings system development. We introduce an easy-to-implement deep-learning-

based model for convective core prediction in this article. NetNCC is developed to help operational forecasters in nowcasting deep convection and potential extreme precipitation in regions with limited ground-based radars. Our model uses satellite observations of TIR T_b as input to probabilistically predict the propagation and initiation of convective cores at short time ranges. NetNCC can be run efficiently on a basic CPU with minimal data processing or data storage needs and provide useful nowcasts at regional scales.

NetNCC uses a neighbourhood-based spatially enhanced loss function that accounts for TIR T_b s in nearby pixels when computing the loss for each pixel in the domain image. FSS is used as the training loss function. The model is tested using a range of metrics in space and time (including FSS at different spatial scales). There is no transfer of error, as the latest nowcasts are available at every 15 min independent of the previous nowcast. Any changes in the movement of cold clouds is captured by the updated nowcast. The model output is updated with new images every 15 min and provides operational forecasters the most up-to-date predictions in near real time. Operational nowcasts for the three domains discussed in this article are available at the UK Centre for Ecology and Hydrology's nowcasting portal <https://africa-hydrology.ceh.ac.uk/nowcasting/>.

The model is tested for three regions: WA, EA, and SA. Skilful predictions are made for a lead time of 1 hr at a very fine spatial resolution (~ 3 km) for all regions. For a lead time of 2 hr, the nowcasts are considered skilful at a spatial resolution of $< \sim 30$ km, whereas for lead time of 4 hr the nowcasts cross the skilful prediction threshold at a spatial resolution of $< \sim 180$ km in SA and EA and at ~ 270 km in WA. Nowcasting accuracy usually changes with the time of day. Prediction accuracy starts improving around 0800 local time for SA and EA as higher FSS values are observed for all lead times during the daytime. In general, the EA domain has relatively higher pixel-wise and spatial-scale-based accuracy at all lead times (particularly at shorter lead times). For a longer lead time of 6 hr, most metrics indicate similar performance of NetNCC nowcasts in the EA and SA domain, with slightly better performance in the SA region at finer spatial scales. WA nowcasting performance, though objectively useful, is relatively lower than the other domains at all lead times for the neighbourhood-based FSS evaluation, whereas it is comparable to the other domains at the pixel-wise level. Compared with persistence, the FSS values show improved accuracy of the nowcasts for all domains at all lead times (in particular for lead times < 3 hr).

Nowcast accuracy is noted to be better for regions where the frequency of core occurrence is higher due to local topographic and convection-enhancing features.

Locations near large water bodies and elevation gradients have relatively higher predictive accuracy, indicating the model's capability to learn the atmospheric contributors to the type of convection generally prevalent in the study areas. The real-time case study from a testbed discussed in this article shows the model's ability to capture the propagation of the convective system as it moves inland from the coastline of Mozambique. NetNCC was able to nowcast a convective system originating from the inland propagation of a sea breeze circulation.

Spatial predictive accuracy decreases with lead time for all regions. Nowcasting at longer lead times may require information about other atmospheric factors as suggested by the model performance. We plan to further explore the utility of the model at longer lead times by adding multiple, independent sources of information at fine and coarse scales regarding land surface and atmospheric conditions. In addition, modifying the model to include information about other non-local features in the form of latent variables could potentially improve the prediction accuracy of convective cores at longer lead times.

ACKNOWLEDGMENTS

We thank Professor Douglas Parker and Professor John Marsham for their insightful comments on process-based evaluation of the nowcasting model. We also extend our gratitude to all the forecasters from the Zambia Meteorological Department, the Instituto Nacional de Meteorologia Mozambique, and the Kenya Meteorological Department who provided valuable suggestions about improving NetNCC during the WISER Testbed. Computation resources were provided by the UK Centre for Ecology and Hydrology.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in MSG High Rate SEVIRI level 1 data at <https://data.eumetsat.int/product/EO:EUM:DAT:MSG:HRSEVIRI>. These data were derived from the following resources available in the public domain: EUMETSAT, <http://www.eumetsat.int>.

A sample ipython notebook outlining the training and testing method for NetNCC is available upon request. Please contact the authors for a PyTorch-based implementation of the code.

ORCID

Jawairia A. Ahmad  <https://orcid.org/0000-0003-2953-9360>

Cornelia Klein  <https://orcid.org/0000-0001-6686-0458>

Christopher Taylor  <https://orcid.org/0000-0002-0120-3198>

Seonaid R. Anderson  <https://orcid.org/0000-0001-8556-577X>

Steven C. Wells  <https://orcid.org/0000-0002-1723-5987>

David C. Hogg  <https://orcid.org/0000-0002-6125-9564>

REFERENCES

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C. et al. (2015) TensorFlow: Large-scale machine learning on heterogeneous systems. <https://www.tensorflow.org/>
- Ai, Y., Li, J., Shi, W., Schmit, T.J., Cao, C. & Li, W. (2017) Deep convective cloud characterizations from both broadband imager and hyperspectral infrared sounder measurements. *Journal of Geophysical Research: Atmospheres*, 122, 1700–1712.
- Alfaro, D.A. & Khairoutdinov, M. (2015) Thermodynamic constraints on the morphology of simulated midlatitude squall lines. *Journal of the Atmospheric Sciences*, 72, 3116–3137.
- Allen, M.R. & Ingram, W.J. (2002) Constraints on future changes in climate and the hydrologic cycle. *Nature*, 419, 224–232.
- Aminou, D. M. (2002) MSG's SEVIRI instrument. *ESA Bulletin* (0376-4265), (111), 15–17.
- Anderson, S.R., Cole, S.J., Klein, C., Taylor, C.M., Diop, C.A. & Kamara, M. (2023) Nowcasting convective activity for the Sahel: a simple probabilistic approach using real-time and historical satellite data on cloud-top temperature. *Quarterly Journal of the Royal Meteorological Society*, 150, 597–617.
- Barton, E.J., Klein, C., Taylor, C.M., Marsham, J., Parker, D.J., Maybee, B. et al. (2025) Soil moisture gradients strengthen mesoscale convective systems by increasing wind shear. *Nature Geoscience*, 18(4), 330–336.
- Bedka, K.M. & Khlopenkov, K. (2016) A probabilistic multispectral pattern recognition method for detection of overshooting cloud tops using passive satellite imager observations. *Journal of Applied Meteorology and Climatology*, 55, 1983–2005.
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X. & Tian, Q. (2023) Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619, 533–538.
- Brier, G.W. (1950) Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78, 1–3.
- Burton, R., Blyth, A., Cui, Z., Groves, J., Lamptey, B., Fletcher, J. et al. (2022) Satellite-based nowcasting of west African mesoscale storms has skill at up to 4-h lead time. *Weather and Forecasting*, 37, 445–455.
- Eigen, D. & Fergus, R. (2015) Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: *2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile*, pp. 2650–2658. Available from: <https://doi.org/10.1109/ICCV.2015.304>
- EUMETSAT. (2025) MSG High Rate SEVIRI level 1.5 data guide. <https://user.eumetsat.int/resources/user-guides/msg-high-rate-seviri-level-1-5-data-guide> [Accessed 25th March 2025].
- Feng, Z., Dong, X., Xi, B., Schumacher, C., Minnis, P. & Khaiyer, M. (2011) Top-of-atmosphere radiation budget of convective core/s-tratiform rain and anvil clouds from deep convective systems. *Journal of Geophysical Research: Atmospheres*, 116, n/a.
- Han, L., Sun, J. & Zhang, W. (2019) Convolutional neural network for convective storm nowcasting using 3-D doppler weather radar data. *IEEE Transactions on Geoscience and Remote Sensing*, 58, 1487–1495.
- Heim, N. & Avery, J.E. (2019) Adaptive anomaly detection in chaotic time series with a spatially aware echo state network. arXiv preprint arXiv:1909.01709.
- Hendon, H.H. & Woodberry, K. (1993) The diurnal cycle of tropical convection. *Journal of Geophysical Research: Atmospheres*, 98, 16623–16637.
- Imhoff, R., Brauer, C., Overeem, A., Weerts, A. & Uijlenhoet, R. (2020) Spatial and temporal evaluation of radar rainfall nowcasting techniques on 1,533 events. *Water Resources Research*, 56, e2019WR026723.
- Kaparakis, C. & Mehrkanon, S. (2023) Wf-unet: weather data fusion using 3D-unet for precipitation nowcasting. *Procedia Computer Science*, 222, 223–232.
- Klein, C. & Taylor, C.M. (2020) Dry soils can intensify mesoscale convective systems. *Proceedings of the National Academy of Sciences*, 117, 21132–21137.
- Klein, C., Belušić, D. & Taylor, C.M. (2018) Wavelet scale analysis of mesoscale convective systems for detecting deep convection from infrared imagery. *Journal of Geophysical Research: Atmospheres*, 123, 3035–3050.
- Lagerquist, R. & Ebert-Uphoff, I. (2022) Can we integrate spatial verification methods into neural network loss functions for atmospheric science? *Artificial Intelligence for the Earth Systems*, 1, e220021.
- Lagerquist, R., Stewart, J.Q., Ebert-Uphoff, I. & Kumler, C. (2021) Using deep learning to nowcast the spatial coverage of convection from Himawari-8 satellite data. *Monthly Weather Review*, 149, 3897–3921.
- Lahuerta, J.A. (2021) Algorithm theoretical basis document for the precipitation product processors of the nwc/geo. Tech. Report. EUMETSAT NWC SAF. https://www.nwcsaf.org/Downloads/GEO/2018/Documents/Scientific_Docs/NWC-CDOP2-GEO-AEMET-SCI-ATBDPrecipitation_v2.1.pdf NWC/GEO ATBD, Version 2.1.
- Lakshmanan, V. & Smith, T. (2010) An objective method of evaluating and devising storm-tracking algorithms. *Weather and Forecasting*, 25, 701–709.
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F. et al. (2023) Learning skillful medium-range global weather forecasting. *Science*, 382, 1416–1421.
- LeCun, Y. & Bengio, Y. (1998) Convolutional networks for images, speech, and time series. In: *The handbook of brain theory and neural networks*. MIT Press. Available from: <https://doi.org/10.5555/303568.303704>
- Lehmann, J. & Coumou, D. (2015) The influence of mid-latitude storm tracks on hot, cold, dry and wet extremes. *Scientific Reports*, 5, 17491.
- Li, J., Zheng, J., Li, B., Min, M., Liu, Y., Liu, C.-Y. et al. (2024) Quantitative applications of weather satellite data for nowcasting: Progress and challenges. *Journal of Meteorological Research*, 38, 399–413.
- Lin, B., Wielicki, B.A., Minnis, P., Chambers, L., Xu, K.-M., Hu, Y. et al. (2006) The effect of environmental conditions on tropical deep convective systems observed from the TRMM satellite. *Journal of Climate*, 19, 5745–5761.
- Machado, L., Rossow, W., Guedes, R. & Walker, A. (1998) Life cycle variations of mesoscale convective systems over the americas. *Monthly Weather Review*, 126, 1630–1654.

- Maqtan, R., Othman, F., Wan Jaafar, W.Z., Sherif, M. & El-Shafie, A. (2022) A scoping review of flash floods in Malaysia: current status and the way forward. *Natural Hazards*, 114, 2387–2416.
- Marr, D. & Hildreth, E. (1980) Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207, 187–217.
- Maybee, B., Marsham, J.H., Klein, C.M., Parker, D.J., Barton, E.J., Taylor, C.M. et al. (2024) Wind shear effects in convection-permitting models influence mcs rainfall and forcing of tropical circulation. *Geophysical Research Letters*, 51, e2024GL110119.
- Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T.R., El-Kadi, A., Masters, D. et al. (2025) Probabilistic weather forecasting with machine learning. *Nature*, 637, 84–90.
- Pulkkinen, S., Nerini, D., Pérez Hortal, A.A., Velasco-Forero, C., Seed, A., Germann, U. et al. (2019) Pysteps: an open-source python library for probabilistic precipitation nowcasting (v1. 0). *Geoscientific Model Development*, 12, 4185–4219.
- Roberts, N.M. & Lean, H.W. (2008) Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 136, 78–97.
- Roberts, C.D., Senan, R., Molteni, F., Boussetta, S., Mayer, M. & Keeley, S.P. (2018) Climate model configurations of the ECMWF integrated forecasting system (ECMWF-IFS cycle 43r1) for High-ResMIP. *Geoscientific Model Development*, 11, 3681–3712.
- Ronneberger, O., Fischer, P. & Brox, T. (2015) U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W. & Frangi, A. (Eds) *Medical image computing and computer-assisted intervention—MICCAI 2015*. MICCAI 2015. Lecture Notes in Computer Science, vol. 9351. Cham: Springer, pp. 234–241. Available from: https://doi.org/10.1007/978-3-319-24574-4_28
- Šakić Trogrlić, R., Van Den Homberg, M., Budimir, M., McQuistan, C., Sneddon, A. & Golding, B. (2022) Early warning systems and their role in disaster risk reduction. In: *Towards the “perfect” weather warning: bridging disciplinary gaps through partnership and communication*. Cham: Springer International Publishing, pp. 11–46.
- Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K. & Woo, W.-c. (2015) Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. and Garnett, R. (Eds.) *Advances in Neural Information Processing Systems*, vol. 28. Curran Associates, Inc. Available from: https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf
- Skok, G. & Roberts, N. (2016) Analysis of fractions skill score properties for random precipitation fields and ECMWF forecasts. *Quarterly Journal of the Royal Meteorological Society*, 142, 2599–2610.
- Stengel, K., Glaws, A., Hettinger, D. & King, R.N. (2020) Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117, 16805–16815.
- Taylor, C.M., Klein, C., Dione, C., Parker, D.J., Marsham, J., Diop, C.A. et al. (2022) Nowcasting tracks of severe convective storms in West Africa from observations of land surface state. *Environmental Research Letters*, 17, 34016.
- Trebing, K., Stańczyk, T. & Mehrkanoon, S. (2021) SmaAt-UNet: precipitation nowcasting using a small attention-UNet architecture. *Pattern Recognition Letters*, 145, 178–186.
- Tuel, A. & Martius, O. (2023) Weather persistence on sub-seasonal to seasonal timescales: a methodological review. *Earth System Dynamics*, 14, 955–987.
- UK Met Office. (2025a) Intensive weather forecast testbed kicks off in Mozambique, South Africa, and Zambia. <https://www.metoffice.gov.uk/services/government/international-development/weather-and-climate-information-services-wiser/wiser-africa/wiser-ews/intensive-weather-forecast-testbed-kicks-off-in-mozambique-south-africa-and-zambia>. [Accessed 27th May 2025].
- UK Met Office. (2025b) Wiser africa. <https://www.metoffice.gov.uk/services/government/international-development/weather-and-climate-information-services-wiser/wiser-africa/index> [Accessed 8th October 2025].
- Weigel, A.P., Liniger, M.A. & Appenzeller, C. (2007) The discrete brier and ranked probability skill scores. *Monthly Weather Review*, 135, 118–124.
- Westra, S., Fowler, H.J., Evans, J.P., Alexander, L.V., Berg, P., Johnson, F. et al. (2014) Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics*, 52, 522–555.
- Yang, G.-Y. & Slingo, J. (2001) The diurnal cycle in the tropics. *Monthly Weather Review*, 129, 784–801.
- Zhao, M. (2014) An investigation of the connections among convection, clouds, and climate sensitivity in a global climate model. *Journal of Climate*, 27, 1845–1862.
- Zipser, E.J., Cecil, D.J., Liu, C., Nesbitt, S.W. & Yorty, D.P. (2006) Where are the most intense thunderstorms on earth? *Bulletin of the American Meteorological Society*, 87, 1057–1072.

How to cite this article: Ahmad, J.A., Klein, C., Taylor, C., Anderson, S.R., Wells, S.C. & Hogg, D.C. (2026) Nowcasting convective cores using deep learning. *Quarterly Journal of the Royal Meteorological Society*, e70284. Available from: <https://doi.org/10.1002/qj.70284>

APPENDIX A. MODEL DEVELOPMENT

Extensive systematic tests were conducted to obtain the current NetNCC framework. The goal was to have a nowcasting model with the smallest number of parameters that require training with the highest achievable prediction accuracy. Initial tests used binary core data as input (similar to NFLICS) instead of TIR T_b to predict binary core masks at the prescribed lead time. However, we found that using normalized TIR as input yielded better accuracy compared with using cores at preceding time steps as input. Successive TIR images contain information about cloud cooling or warming and can, therefore, better predict core expansion or disintegration, particularly at smaller spatial scales. The models trained for specific domains (and specific months) were tested using input data from

another domain (and different months) to evaluate model transferability skills. The results showed that the models were able to predict the general location of core occurrence with varying probability values, indicating the model's ability to learn the first-order physics behind convective core development. However, the models were often unable to correctly predict the exact location of core occurrence due to different learned core propagation speeds.

Similarly, different versions of the model were trained and tested where one set included the time of day being predicted as input while the other set ignored the impact of time of day in predicting cores at all lead times. It was observed that adding the time of day as a positive sine function (0 at midnight and 1 at noon) improved the prediction accuracy at longer lead times. Preliminary tests included adding time of day as an input channel such that all pixels had the same hour and minute value of the time being predicted. However, that decreased the sensitivity of the nowcast to the TIR image at time t_0 (time when prediction is initialized), and thus decreased the nowcast accuracy. Adding time of day as a variable during the last image reconstruction skip connection (see Figure 1a) worked best in terms of nowcasting accuracy.

An important testing step during model development focused on using near versus far past TIR images to predict the future. Sample tests were conducted using TIR images from time steps in the past equal to the lead time. In other words, for a lead time of 2 hr, three TIR images at time t_0 , $t_0 - 2$ hr, and $t_0 - 4$ hr were used as input. The results showed that for lead times ≥ 1 hr the TIR image at time t_0 and $t_0 - 1$ hr were most important for accurate prediction. Considering the rapidly changing atmospheric conditions during the rainy season within the domains, it is expected that information from more than 2 hr prior to the time of prediction would have limited impact on predicting small-to medium-scale convective activity in the future. The test results corroborated it.

NetNCC uses spatial information to compute the loss function during training, and thus it is expected that the scale of the neighbourhood filter would impact the spatial prediction accuracy. Two sets of simulations tested the sensitivity of the prediction accuracy to the training neighbourhood filter size and the testing spatial scale used to compute the statistical accuracy (FSS). It was seen that the model was less impacted by the training filter size as long as the radius was kept large enough to capture small cores (~ 10 km) but small enough to avoid excessive-freedom issues (which could lead to striped patterns in the predicted images). A training filter window of 5×5 pixels (~ 15 km $\times$ 15 km) gave the best results across different spatial evaluation scales.

The prediction accuracy is considerably impacted by the evaluation filter window (in other words, spatial scale)

at which the statistical accuracy (FSS) is computed. There is a positive, nonlinear relationship between the spatial scale used to compute FSS and the FSS value. This indicates that the nowcasting accuracy takes into account the occurrence of cores in the neighbourhood. Thus, even if the location of the nowcasted core is not precisely predicted, the FSS value considers the occurrence of the core in the general neighbourhood defined by the evaluation filter size. Considering the relationship between extreme rainfall and convective cores, it is important that the nowcasting model correctly predicts the presence of extreme rainfall inducing convective systems within the general neighbourhood. Thus, the increase in accuracy with respect to spatial scale bodes well for impact-based nowcasting.

APPENDIX B. DIURNAL CYCLE IN MODEL ACCURACY

We provide separate accuracy maps for lead times of 2, 4, and 6 hr at different times of day for the three different domains to further explain the nowcasting ability of the model (Figures B1, B2, and B3). The errors are quite small across the domains for lead time of 1 hr when the nowcasts are most reliable and have thus been excluded. These maps highlight the relationship between nowcast accuracy and the diurnal cycle for each lead time. Operational forecasters can then refer to these maps to get an estimate of the nowcast reliability at each location at a particular time of day. Static features, such as elevation gradients and large water bodies, physically constrain convection in the SA (Figure B1) and EA (Figure B2) domains, whereas sea breeze effects are a prominent control on convection in the WA (Figure B3) domain. The Brier score values for all domains are quite low but have consistent spatial patterns.

The diurnal cycle in convective initiation and propagation is quite noticeable in the figures. Figure B1 shows higher Brier score values in SA from midnight to noon local time over regional lakes, in particular Lake Malawi (13° S, 34° E), with increasing magnitudes as the lead time goes from 2 to 6 hr. This corresponds well with the general location of convective cores during the night when land breeze moves towards the warmer water bodies. From noon to midnight the process is reversed, and we note higher Brier score values along elevation gradients that serve as static features driving the wind upwards. The largest Brier score values are noted from noon to 1800 h local time, when the majority of cores occur. It is difficult to predict in the afternoon (when the atmosphere is quite unstable) what the conditions would be in the evening, particularly when nowcasting for 6 hr ahead. We see this manifested in the relatively larger errors in nowcasts for 1200 h–1800 h local time at a lead time of 6 hr

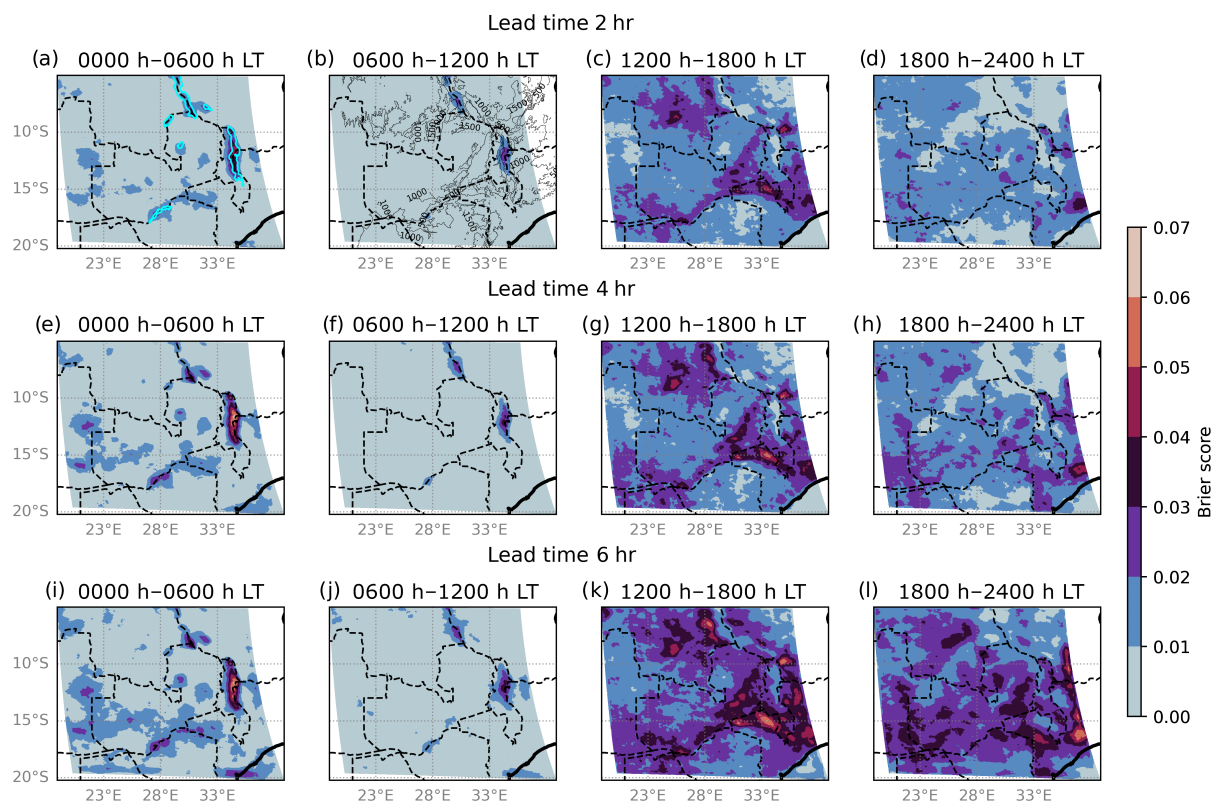


FIGURE B1 Brier score computed using nowcasts for December–February at ~45 km for lead times 2, 4, and 6 hr at different predicted times of day (LT: local time) for the southern Africa domain. Lakes in (a) are outlined in cyan, and the elevation contours in (b) are shown in grey.

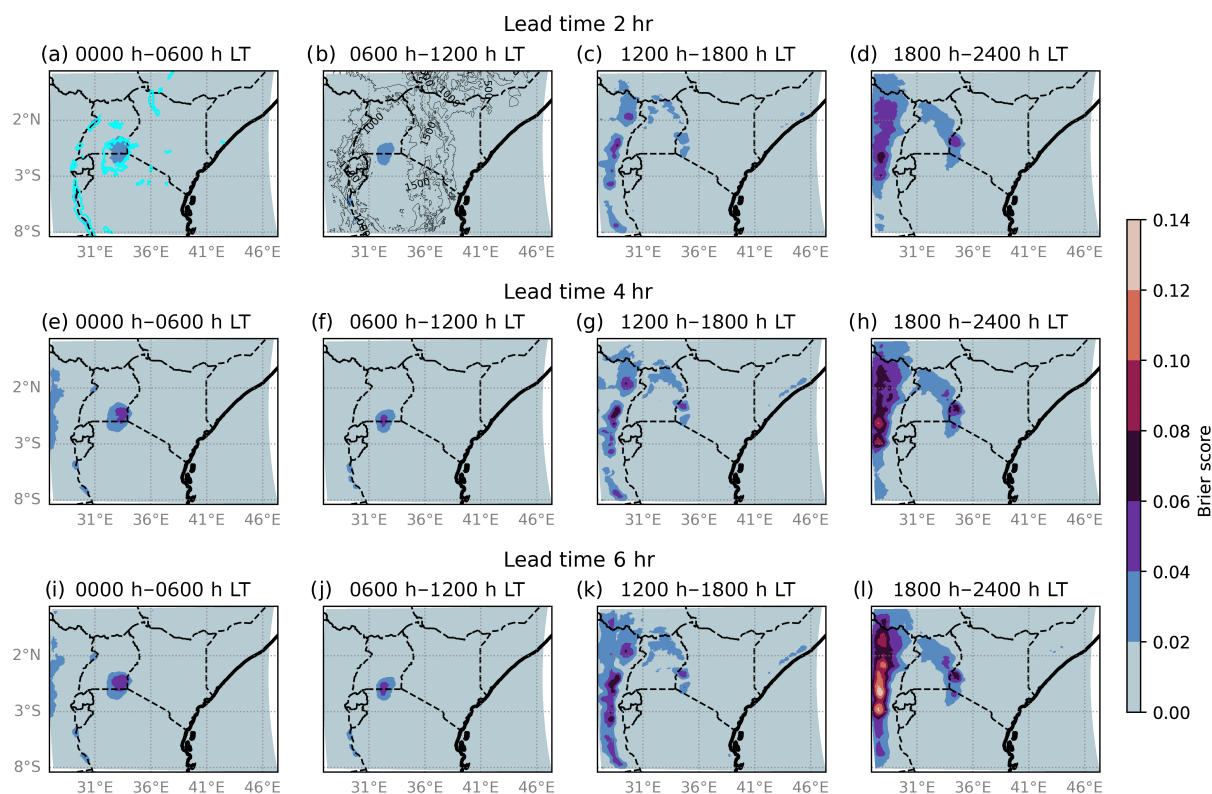


FIGURE B2 As Figure B1 but for March–May in the eastern Africa domain. Note that the colour bar scale is different here.

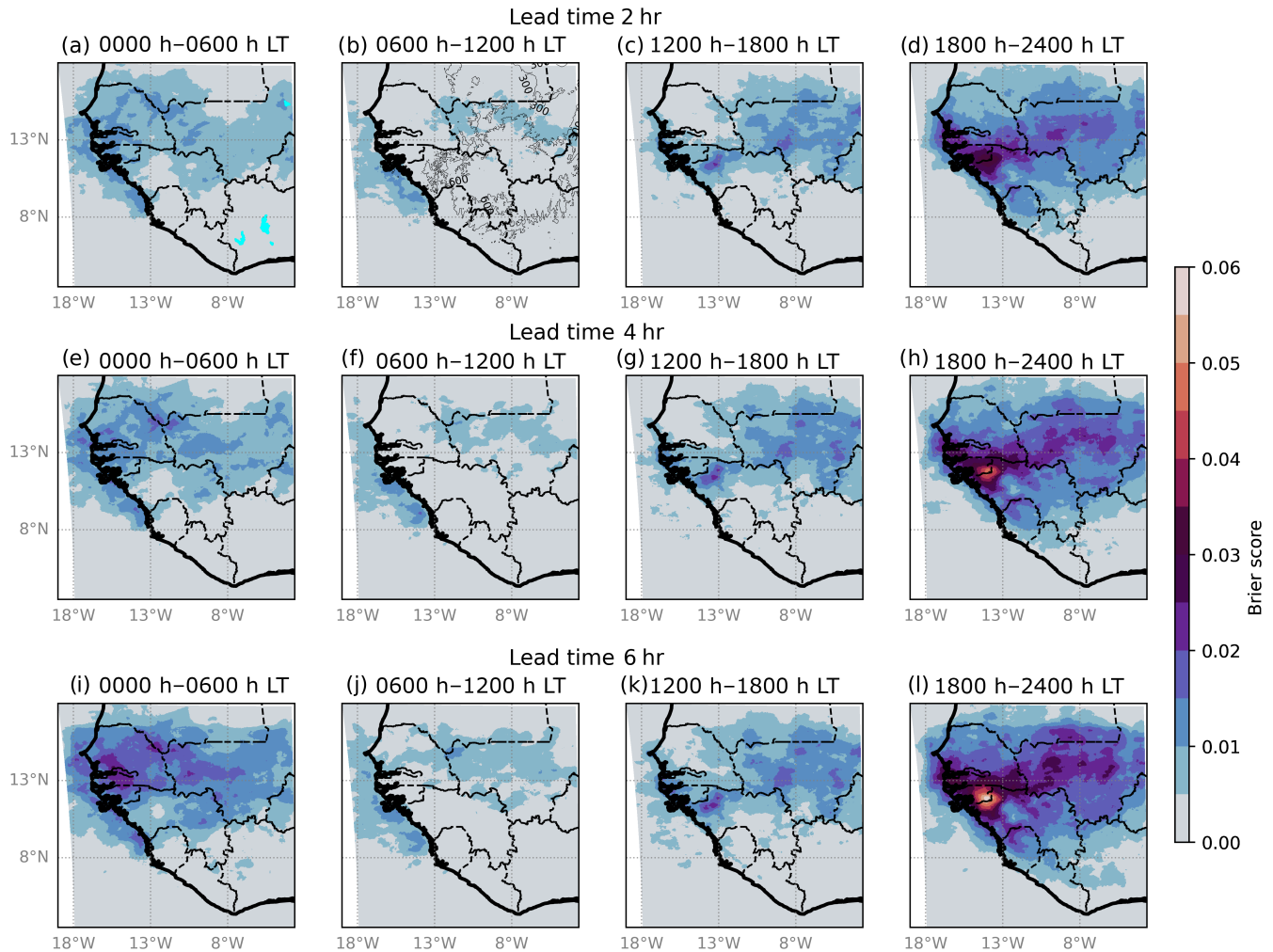


FIGURE B3 As Figure B1 but for July–September in the western Africa domain. Note that the colour bar scale is different here.

(Figure B1k). Similar results are apparent in Figure B2 for EA as well, where we see higher Brier scores over Lake Victoria (1° S, 33° E) for nowcasts from midnight to noon. From noon to midnight, higher Brier scores are seen along the high-elevation gradients on the borders between the Democratic Republic of the Congo and Uganda, Rwanda, Burundi, and Tanzania. The maximum Brier score value is higher here than the SA or WA domains as the errors are aggregated over frequently occurring convective cores along the mountains.

There is relatively limited influence of inland water bodies in the WA domain (Figure B2). Land breeze effects seem to be more important from midnight to noon, when higher Brier score values are seen along the coast of Senegal and downwards. However, higher Brier score values are noted along elevation gradients around 11° N, 13° W for nowcasts from noon to midnight. Westwards-propagating mesoscale convective systems dominate the nowcasts in WA, as visible from the higher Brier score values from 1800 h to midnight local time. The spatial patterns in Brier score values during the diurnal cycle are similar

for different lead times, albeit with increasing magnitudes from 2 to 6 hr lead time. However, all of the Brier score values are objectively quite low, with the maximum values less than 0.2. The model derives its skill from spatio-temporal TIR information inherent in the input maps that identify cold regions where existing convective systems are present and whether the existing systems have grown or disintegrated over the last couple of hours based on cooling or warming T_b values. In addition, newly developing systems are highlighted by cold T_b values appearing in the last hour at locations that previously had no sub-zero temperature values. The model's skill is dependent (by varying degrees) on the learned convective core climatology and the recent convection-related information inherent in the TIR T_b inputs based on the lead time. Recent TIR T_b information is the primary source of skill for both existing and newly developing convective systems at shorter lead times, whereas the learned climatology becomes more important in predicting the direction and speed of propagation of existing cores at longer lead times.