Article:

Rakotomanga, M., Parker, D.J., Ben Rached, N. et al. (Accepted: 2026) Object-Based Deep Learning for Probabilistic Convective Core Nowcasting from Satellite Data. Quarterly Journal of the Royal Meteorological Society. ISSN: 0035-9009 (In Press)

OBJECT-BASED DEEP LEARNING FOR PROBABILISTIC CONVECTIVE CORE NOWCASTING FROM SATELLITE DATA

Mendrika Rakotomanga[1,*], **Douglas J. Parker**[1,2,3], **Nadhir Ben Rached**[1], **Seonaid R. Anderson**[4], **Cornelia Klein**[4], **Steven C. Wells**[4]

[1]School of Mathematics, University of Leeds, UK
[2]School of Earth and Environment, University of Leeds, UK
[3]NORCE Norwegian Research Centre AS, Bergen, Norway
[4]UK Centre for Ecology & Hydrology, Wallingford, UK
[*]Corresponding author: mmmhr@leeds.ac.uk

ABSTRACT

Flash flooding from intense rainfall causes major damage and loss of life across Africa, particularly in the Sahel, where rainfall is dominated by mesoscale convective systems. Convective cores, which represent regions of intense convective activity, evolve rapidly, limiting short-term predictability, especially in data-sparse regions where numerical weather prediction models face substantial uncertainty in convective initiation, evolution, and organisation. This study presents an object-based deep learning framework for probabilistic convective core nowcasting using geostationary satellite data. Convective cores are identified from Meteosat Second Generation infrared imagery using a two-dimensional wavelet transform applied at mesoscale spatial scales. The framework is first demonstrated at a single location, where nearby core attributes are used to predict local core occurrence up to six hours ahead. Explainable AI analysis indicates that predictions are driven by physically meaningful factors related to core proximity, size, and intensity. The approach is then extended to the western Sahel through NCAST, a spatio-temporal Transformer architecture that models the evolution and interactions of convective core populations and produces gridded probabilistic nowcasts. Ablation experiments reveal that self-attention contributes most strongly to forecast skill, while temporal information becomes increasingly important at longer lead times. Evaluation against persistence and an operational conditional-climatology benchmark demonstrates skilful forecasts at 1-, 3-, and 6-hour lead times using multiple probabilistic and spatial verification metrics. These results highlight the potential of object-based learning to provide reliable short-range nowcasts in data-sparse environments.

Keywords

Nowcasting; Satellite observations; Convective storms; Deep learning; Object-based methods; Probabilistic forecasting

1 Introduction

Mesoscale convective systems (MCSs) are large, long-lived clusters of thunderstorms that occur throughout the tropics, mid-latitudes, and subtropics (Laing & Fritsch, 2000; Zipser et al., 2006). In the Sahel, they provide a major source of rainfall and are frequently associated with flash floods, damaging winds, intense lightning, and other high-impact hazards (Mathon et al., 2002; Panthou et al., 2018; Klein et al., 2021). Convective cores, which correspond to the most active and coldest parts of convective systems, are short-lived and highly dynamic (Markowski & Richardson, 2010; Klein et al., 2018). Rapid evolution, strong diurnal variability, and sensitivity to environmental triggers make these cores inherently difficult to forecast, particularly in tropical regions such as Africa, where sparse in situ observations further constrain short-range predictability associated with convective dynamics (Vogel et al., 2018; Kniffka et al., 2020; Keane et al., 2025).

Nowcasting, defined by World Meteorological Organization (2017) as forecasting with local detail up to six hours ahead, provides a practical framework for anticipating rapidly evolving convection. Radar networks across Africa remain sparse (Gijben & De Coning, 2017; World Meteorological Organization, 2025), making geostationary satellite observations, including Meteosat Second Generation (MSG), essential. Recent satellite-based nowcasting systems include motion-based approaches, such as cell tracking and optical flow, which typically provide useful skill for short-term extrapolation of convective systems and associated rainfall up to 1–4 hours ahead (Pulkkinen et al., 2019; Hill et al., 2020; Burton et al., 2022). To address the limitations of pure extrapolation, hybrid schemes combine cloud-top and land-surface information to capture both propagation and initiation (Taylor et al., 2022).

Machine learning methods have increasingly been applied to modelling convective evolution. Early studies applied classical algorithms to storm detection, hazard classification, and short-term evolution using radar-derived predictors (Kitzmiller et al., 1995; Billet et al., 1997; Gagne et al., 2009; Mecikalski et al., 2021). More recently, deep learning has enabled richer spatial and temporal representations: convolutional neural networks extract cloud-top structures (Cintineo et al., 2020), recurrent networks model temporal evolution (Leinonen et al., 2023), encoder–decoder architectures generate pixel-level nowcasts of convective proxies (Lagerquist et al., 2021), and generative models produce probabilistic storm nowcasts from satellite imagery (Dai et al., 2025; Hermes et al., 2025). However, most of these systems rely on image-based prediction and have been developed primarily outside Africa, where observational density and operational constraints differ markedly (Roberts et al., 2022; Taylor et al., 2022; Anderson et al., 2024).

In contrast, object-based approaches treat convection as a set of discrete features rather than continuous images. Such representations reduce data dimensionality, improve interpretability, and align closely with physical storm processes (Dixon & Wiener, 1993; Handwerker, 2002). Anderson et al. (2024) introduced an object-based conditional climatology for the Sahel, providing domain-wide probabilistic nowcasts of convective-core occurrence based on the current locations of observed cores. While computationally efficient and operationally useful, its predictions depend on the choice of neighbourhood scale and do not explicitly incorporate core attributes such as size, intensity, or proximity, which influence convective evolution. By conditioning solely on spatial location through historical neighbourhood statistics, the approach does not explicitly represent convective initiation, interaction, or decay, which may limit its ability to capture event-to-event variability in convective structure and activity.

These limitations motivate investigation into how far the short-term evolution of convective cores can be inferred from object-level information alone. In this study, we develop an object-based deep learning framework trained on compact physical descriptors of convective cores derived from geostationary satellite observations at mesoscale resolution. We first assess local predictability, examining whether the properties of nearby cores, including their relative location, size, and intensity, contain sufficient information to forecast local core occurrence. We then extend the framework to the western Sahel by representing the population of cores over a two-hour history and modelling their evolution using a spatio-temporal Transformer coupled with a convolutional decoder. Together, these experiments assess the extent to which physically informed, low-dimensional descriptors can support probabilistic nowcasting in data-sparse environments.

The remainder of this paper is organised as follows. Section 2 describes the convective core dataset and the conditional-climatology benchmark. Section 3 presents the object-based modelling framework for single-location and spatial nowcasting. Section 4 presents the evaluation methodology and results. Section 5 examines the physical interpretation, operational considerations, and limitations of the approach. Section 6 concludes the paper.

2 Data sources and nowcast benchmark

2.1 Cloud-top temperature and convective core data

This study focuses on the June–September monsoon season and uses the western Sahel as the primary domain (Figure 1). The region spans roughly 5°–20° N and 19°–5° W, covering Senegal, Mauritania, Mali, and western Niger, where deep convection is frequent during the monsoon (Laing & Fritsch, 1993; Zipser et al., 2006; Taylor et al., 2017; Maranan et al., 2018). Cloud-top temperature (CTT) derived from MSG infrared (IR) observations (EUMETSAT, 2022) provides an effective basis for identifying convective structures. Colder cloud tops correspond to higher cloud altitudes and are strongly associated with deep convection (Zipser et al., 2006; Adler & Mack, 1986). Although CTT does not measure rainfall directly, it exhibits strong statistical relationships with heavy rainfall, lightning, hail, and damaging winds (Maidment et al., 2014). For this study, we use the 2004–2024 archive of MSG SEVIRI IR 10.8 µm brightness temperatures at 15-minute temporal resolution and approximately 3–4 km spatial resolution over Africa. These observations provide a long, consistent record that has been widely used to monitor convective activity across the continent (Taylor et al., 2017; Anderson et al., 2024; Hermes et al., 2025).

Convective cores are identified using the method described by Anderson et al. (2024), which builds upon the wavelet-based approach of Klein et al. (2018) to detect localised cold features within CTT fields using a multiscale two-dimensional wavelet transform. Pixels colder than $-40°C$ are first retained to isolate deep cloud regions associated with convection. A Marr wavelet (Marr & Hildreth, 1980) is applied across spatial scales spanning approximately 12–150 km. Wavelet power is computed across the range of spatial scales and summed, with candidate features identified where the resulting summed wavelet power exceeds the scale-dependent threshold $(a + f)^{0.5}$, where a denotes the wavelet scale and $f = -8$. Features whose minimum CTT is warmer than $-50°C$ are excluded to suppress cloud-edge artefacts. For each retained feature, the position $\mathbf{s}_0$ of maximum wavelet power is recorded as the convective core location. While the underlying wavelet methodology was originally developed in the context of MCS analysis (Klein et al., 2018), the identified features are interpreted here as individual convective cores and are not restricted to MCSs.

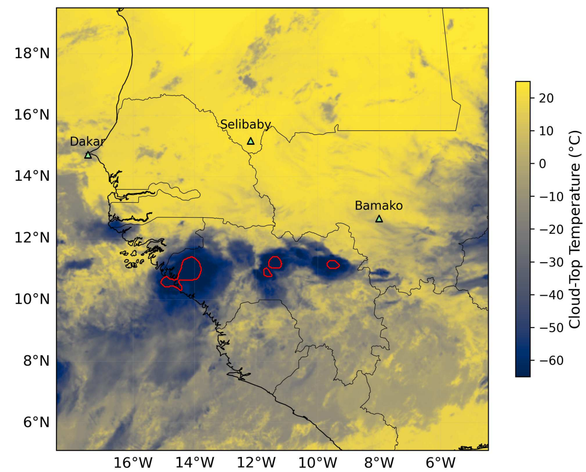


Figure 1: Cloud-top temperature (CTT; °C) from the MSG IR 10.8 µm channel over the western Sahel study domain on 3 July 2023 at 1515 UTC. Convective cores (red outlines) are identified using a two-dimensional wavelet transform.

2.2 Conditional climatology benchmark for probabilistic convective core nowcasting

As a reference system, we use the conditional climatology nowcasting approach introduced by Anderson et al. (2024). This method produces gridded probabilistic nowcasts based on the observed locations of convective cores at the nowcast origin time t_0.

The unconditional climatology is constructed by averaging historical binary convective-core fields at each time of day t using data from 2004–2019. The binary field $\mathbf{C}_t$ assigns value 1 to grid cells associated with a detected convective core (wavelet power > 0) and 0 otherwise. Each field is smoothed using a maximum neighbourhood filter of size L, ranging from 15 to 155 km, yielding the unconditional probability field $P(\mathbf{C}_t)$.

The conditional climatology restricts the historical sample to dates on which an identified convective core was present within a predefined neighbourhood S_0 surrounding the observed core location $\mathbf{s}_0$ at time t_0. For these dates, the neighbourhood-filtered fields at verification times $t = t_0 + \tau$, where τ denotes the forecast lead time, are averaged to produce the conditional probability field $P(\mathbf{C}_{t_0+\tau} \mid \mathbf{s}_0)$. When multiple cores are present at t_0, conditional probability fields are computed independently for each core and combined using a pixel-wise maximum.

Throughout this paper, we refer to this operational implementation as NFLICS (Nowcasting Flood Impacts of Convective Storms in the Sahel), the project under which the system was developed and deployed (UKRI Gateway to Research, 2025). Although originally developed for the western Sahel, it now operates across Africa in real time via the UKCEH nowcasting portal (UK Centre for Ecology & Hydrology, 2024). Verification using 2020–2022 demonstrates skilful probabilistic forecasts at 1–6 h lead times, with forecast skill dependent on neighbourhood scale while remaining computationally lightweight (Anderson et al., 2024).

While NFLICS has demonstrated skill and operational utility, its formulation reflects specific structural design choices. Because the method conditions solely on core locations at t_0, it does not explicitly represent recent motion, life-cycle stage, or interactions between nearby convective features. In addition, the pixel-wise maximum used to combine conditional fields when multiple cores are present can yield spatially extensive regions of non-zero probability. These characteristics motivate exploration of complementary object-based learning approaches that incorporate additional core attributes and short-term temporal context.

3 Methodology

We develop two deep learning models for convective core nowcasting, both built on a physically motivated object-based representation of convective cores extracted from MSG retrievals. The objective is to assess the predictability captured by core-level information alone and whether this formulation supports both local and spatial forecasting. The first model provides a simplified proof-of-concept application and produces probabilistic forecasts of core occurrence at a single location in Dakar. It is formulated as a binary classification task for forecast lead times $\tau \in \{1, 3, 6\}$ hours, where forecasts issued at time t_0 are verified at $t_0 + \tau$. The second model extends this approach to generate gridded nowcasts across the western Sahel at the same lead times. In both cases, a separate model is trained for each forecast lead time. The two approaches use an identical data split: 2004–2018 for training, 2019 for validation, and 2020–2024 for independent testing. The following subsections describe the object representations, model architectures, and training procedures used in each framework.

3.1 Single-location nowcasting: application to Dakar

The single-location nowcasting framework is designed to predict convective-core occurrence at a specified target site using information from surrounding convective cores at the nowcast origin time t_0. Dakar, with a population exceeding three million, is selected as an illustrative application. As a major urban centre on Senegal's western coast, it is frequently affected by organised convective storm systems (Laing & Fritsch, 1993; Zipser et al., 2006; Agence Nationale de la Statistique et de la Démographie, 2023). Its exposure to propagating West African convection makes it an appropriate testbed for evaluating short-range nowcasting approaches aimed at reducing weather-related risks in densely populated areas.

For this application, the context domain is defined as the region from which convective cores could realistically influence Dakar within a 6-hour period. Based on typical propagation speeds in West Africa, which rarely exceed $100\,\mathrm{km\,h^{-1}}$ (Mathon et al., 2002), this corresponds to an approximately $1300 \times 1300\,\mathrm{km}$ ($\sim 12^\circ \times 12^\circ$) square centred on Dakar. Within this domain, an average of three convective cores are identified per 15-minute analysis time during periods when convective cores are present (Appendix A). In this configuration, only information available at t_0 is used. The resulting feature vector $\mathbf{X}_{t_0}$ comprises core attributes and timestamp information and is used to generate forecasts at lead time τ.

3.1.1 Data preparation

At t_0, the three nearest convective cores to the target location are described by their latitude, longitude, wavelet power, size (defined as the number of grid cells multiplied by $9\,\mathrm{km}^2$ per pixel), and Haversine distance from Dakar. When fewer than three cores are present, the input is padded with artificial non-convective cores (size $= 0$, wavelet power $= 0$) placed randomly just outside the context domain to avoid introducing fixed spatial artefacts. This padding ensures a fixed-length input without distorting the scaled feature distributions.

For each forecast lead time τ, a tabular dataset is constructed in which each row contains the core features together with a binary target variable $y_{t_0+\tau} \in \{0, 1\}$. For Dakar, $y_{t_0+\tau} = 1$ if at least one grid cell with positive wavelet power is present within a 5×5 window (approximately $15\,\mathrm{km} \times 15\,\mathrm{km}$) centred on Dakar at $t_0 + \tau$, and $y_{t_0+\tau} = 0$ otherwise. Table 1 shows an excerpt of the 1-hour training dataset.

year	month	day	hour	minute	lat1	lon1	lat2	lon2	lat3	lon3	wp1	wp2	wp3	size1	size2	size3	d1	d2	d3	y_1h
2016	8	28	16	30	14.71	-17.46	14.33	-17.03	13.88	-17.62	115.58	159.60	93.74	1485	567	243	1.44	63.87	94.44	0
2016	8	28	16	45	14.74	-17.52	14.33	-16.97	13.25	-17.41	101.99	147.14	126.91	1089	990	288	6.12	68.86	163.47	0
2016	9	9	19	15	14.68	-17.39	11.62	-16.95	11.14	-18.86	119.89	129.46	138.46	2628	21582	117	8.99	348.54	425.45	0
2016	9	9	19	30	14.68	-17.45	13.04	-17.21	11.17	-18.98	115.17	125.62	157.18	1791	22590	297	4.13	187.88	427.25	0
2016	9	12	5	0	14.74	-17.43	12.36	-17.34	15.84	-15.31	143.53	101.50	149.89	819	5013	9189	4.96	262.00	262.58	1
2016	9	12	5	15	14.77	-17.49	15.92	-15.44	12.37	-17.43	132.18	138.48	99.04	999	7785	4275	6.40	255.63	261.53	0
2016	9	12	5	30	14.77	-17.52	16.06	-16.71	16.04	-15.63	112.63	220.97	139.21	1026	522	6876	8.29	169.86	246.13	0
2016	9	13	2	15	14.77	-17.49	11.97	-19.62	14.07	-12.62	98.80	100.99	118.60	2628	1287	612	6.40	384.14	527.48	0
2016	9	16	17	15	14.71	-17.40	12.47	-14.46	14.88	-13.17	127.24	133.59	133.20	1008	666	414	7.84	410.11	462.74	0

Table 1: Excerpt from the raw dataset used to train the Dakar convective-core nowcasting model for a 1 h forecast lead time. Indices 1–3 denote the three nearest convective cores to Dakar at the nowcast origin time t_0, ordered by increasing distance. Variables `year`, `month`, `day`, `hour`, and `minute` describe the nowcast origin time. Variables `lat`, `lon`, `wp`, `size`, and `d` denote latitude (°), longitude (°), wavelet power (K^2), core area (km^2), and Haversine distance from Dakar (km), respectively. The target variable `y_1h` corresponds to y_{t_0+1h} and indicates the presence (1) or absence (0) of a convective core over Dakar at the verification time $t_0 + 1h$.

A binary `mask` variable is introduced to indicate whether each core corresponds to a real observation or to an artificially padded one. Together with the physical core attributes and the full timestamp at the nowcast origin t_0, this yields a 23-dimensional input vector $\mathbf{X}_{t_0}$. The variables `size`, `wp` (wavelet power), and `d` (distance) are log-transformed prior to normalisation to reduce skewness and stabilise variance, thereby facilitating optimisation and improving training stability (LeCun et al., 2012).

The nowcast-origin time variables (`year`, `month`, `day`, `hour`, `minute`) are included as integer-valued predictors, providing a simple representation of temporal information in this proof-of-concept framework. Normalisation is applied after all transformations to all features except the mask, using means and standard deviations computed exclusively from real convective cores in the training set. This procedure avoids information leakage and ensures that all input features lie on comparable scales.

3.1.2 Model architecture for a single location

We adopt a multilayer perceptron (MLP) as the primary deep-learning architecture for the single-location nowcasting problem. The MLP provides a simple deep-learning framework against which the benefits of incorporating additional temporal and spatial information can be assessed.

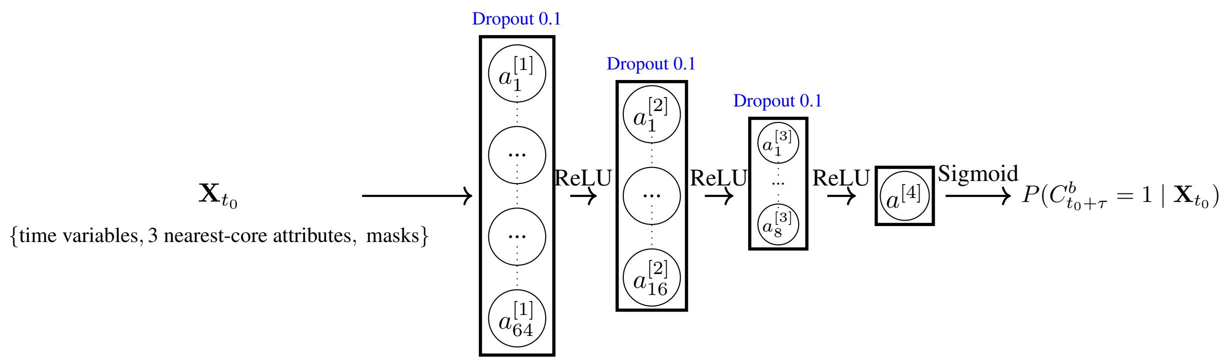


Figure 2: Neural network architecture of the single-location MLP. The illustrated configuration corresponds to the architecture selected for the 3 h forecast lead time. The input vector $\mathbf{X}_{t_0}$ contains time variables, attributes of the three nearest convective cores at the nowcast origin time t_0, and binary mask variables indicating padded cores. The model comprises hidden layers containing 64, 16, and 8 neurons with ReLU activations and dropout regularisation. A sigmoid output neuron produces the probability $P(C^b_{t_0+\tau} = 1 \mid \mathbf{X}_{t_0})$ of convective-core occurrence over the target location. The architectures selected for the other forecast lead times are summarised in Table 2.

Figure 2 provides a schematic illustration of the single-location nowcasting architecture. A separate MLP is trained for each forecast lead time $\tau \in \{1, 3, 6\}$ h to predict convective-core occurrence over the target location. The model takes as input the feature vector $\mathbf{X}_{t_0}$ and consists of a sequence of fully connected hidden layers with ReLU activations (Nair & Hinton, 2010), followed by a single sigmoid output neuron that produces the predicted probability $\hat{y}_{t_0+\tau}$. Dropout is applied between hidden layers to reduce overfitting (Srivastava et al., 2014).

Training minimises the binary cross-entropy (BCE) loss (Bishop, 2006) between $\hat{y}_{t_0+\tau}$ and the ground truth $y_{t_0+\tau}$. Model parameters are optimised using the Adam algorithm (Kingma & Ba, 2015). Hyperparameters, including the hidden-layer configuration, dropout probability, learning rate, and batch size, are selected independently for each forecast lead time using grid search, with the area under the receiver operating characteristic curve (AUC-ROC; Fawcett, 2006) on the validation set used for model selection. During training, early stopping is based on validation AUC with a patience of 10 epochs. The resulting configurations are summarised in Table 2.

To assess whether the proposed deep-learning framework provides advantages over established tabular machine-learning approaches, four benchmark models were also considered: logistic regression (LR; Cox, 1958), support vector machines (SVM; Cortes & Vapnik, 1995), random forests (RF; Breiman, 2001), and Extreme Gradient Boosting (XGBoost; Chen & Guestrin, 2016). These models were trained using the same input representation as the MLP and therefore provide a direct benchmark for evaluating the proposed framework. For each forecast lead time, hyperparameters were selected using grid search with the same selection criterion. The search was performed independently for each model and forecast lead time, and the optimal configurations are summarised in Appendix B.

To assess whether minimal temporal context improves short-range predictability, two additional variants are considered. The first, denoted MLP-2T, extends the single-time-step input representation by concatenating the feature vectors

$\mathbf{X}_{t_{-1}}$ and $\mathbf{X}_{t_0}$, where $t_{-1} = t_0 - 1\,\mathrm{h}$ and $\mathbf{X}_{t_{-1}}$ is constructed identically to $\mathbf{X}_{t_0}$. The concatenated feature vector is then processed by an MLP. The second variant, denoted LSTM-2T, treats the two feature vectors as a short temporal sequence and processes them using a Long Short-Term Memory (LSTM) network (Hochreiter & Schmidhuber, 1997). The final hidden state provides a fixed-length representation of recent convective evolution, which is subsequently passed to fully connected layers to produce the forecast probability.

For each forecast lead time and model variant (MLP, MLP-2T, and LSTM-2T), hyperparameters were selected using grid searches with validation AUC as the selection criterion. For the MLP and MLP-2T models, the search considered dropout probabilities of 0.0, 0.1, 0.2, and 0.3; learning rates of 10^{-3}, 5×10^{-4}, and 10^{-4}; batch sizes of 128 and 256; and hidden-layer configurations of 64–16–8 and 128–64–32–16, yielding 48 candidate configurations per forecast lead time. For LSTM-2T, the search additionally considered LSTM layers with 20 or 32 units and dense-layer configurations of 16 and 32–16, together with dropout probabilities of 0.0 and 0.2, learning rates of 10^{-3} and 10^{-4}, and batch sizes of 128 and 256, yielding 32 candidate configurations per forecast lead time. The best-performing configuration for each case is summarised in Table 2.

Model	Lead time	Hidden layer units	LSTM units	Dropout probability	Learning rate	Batch size
MLP	1 h	128–64–32–16	–	0.2	10^{-3}	128
	3 h	64–16–8	–	0.1	5×10^{-4}	128
	6 h	128–64–32–16	–	0.1	10^{-3}	128
MLP-2T	1 h	128–64–32–16	–	0.2	5×10^{-4}	128
	3 h	128–64–32–16	–	0.3	5×10^{-4}	128
	6 h	128–64–32–16	–	0.3	10^{-3}	256
LSTM-2T	1 h	32–16	32	0.2	10^{-3}	128
	3 h	32–16	20	0.2	10^{-3}	128
	6 h	32–16	32	0.2	10^{-3}	128

Table 2: Best hyperparameter configurations selected using validation AUC for each forecast lead time and model variant. Hidden layer units denote the number of neurons in successive dense layers. All models use a single sigmoid output neuron.

3.2 Nowcasting convective cores for a domain: NCAST

The location-specific formulation is designed to predict convective-core occurrence at a single target location and does not produce spatially continuous forecasts over the analysis domain. To explicitly represent spatial structure and interactions between cores, this framework is extended to a domain-wide nowcasting model over the western Sahel. Detected convective cores are represented as individual tokens characterised by their temporal, physical, and spatial attributes and are processed jointly using a spatio-temporal Transformer encoder. Through self-attention, the model learns interactions between core tokens across both space and time. The resulting token representations are aggregated using mask-weighted mean pooling to obtain a global latent representation, which is projected onto a two-dimensional latent grid and subsequently passed to a convolutional decoder to produce gridded probabilistic nowcasts of convective-core occurrence at forecast lead time τ. This domain-scale architecture is referred to as NCAST (Nowcasting with a Core-Aware Spatio-temporal Transformer).

3.2.1 Data preparation

The input representation follows the same object-based philosophy as the single-location model but is adapted for domain-wide nowcasting. For each time step, the number of detected convective cores is capped at the 99.99th percentile of observed core counts, corresponding to a maximum of $N = 50$ cores per frame (Appendix A). When fewer than 50 cores are present, the remaining entries are padded with dummy cores randomly placed within the domain and assigned zero area and a non-convective CTT of 30°C. When more than 50 cores are detected, the 50 cores with the coldest CTT values are retained.

Each core is represented by a feature vector comprising the latitude and longitude of the pixel of maximum wavelet power; the minimum and maximum latitude and longitude of the bounding box; core area (number of pixels multiplied by 9 km^2); the mean CTT within a 3×3 neighbourhood centred on the coldest pixel; and a binary mask indicating whether the entry corresponds to a real or padded core.

Month and time of day at each lag are encoded using sine–cosine pairs

$$\left(\sin\left(2\pi\frac{\theta}{P}\right), \cos\left(2\pi\frac{\theta}{P}\right)\right),$$

where θ denotes the month index or time-of-day value and P the corresponding period (12 months or 24 hours). These cyclic encodings provide a continuous representation of seasonal and diurnal variability and form part of the initial token representation before embedding by the Transformer encoder. The calendar year is omitted because it does not correspond to a cyclic physical process and may encourage memorisation of interannual variability rather than learning generalisable convective behaviour. The day of month is also excluded, as SHAP analyses of the single-location model indicated low predictive importance (Section 4.1.2). Appending the four temporal features to the nine spatial and physical attributes yields $F = 13$ features per token.

All positional and physical features are standardised using statistics computed exclusively from real cores in the training set. The temporal encodings and binary mask are left unscaled, as they are already bounded and dimensionless.

Inputs are sampled every 30 minutes from $t_0 - 2\,\mathrm{h}$ (t_{-2}) to t_0, producing a spatio-temporal tensor

$$\mathbf{X} \in \mathbb{R}^{T\times N\times F},$$

with $T = 5$ time steps, $N = 50$ cores per time step, and $F = 13$ features per core. For each forecast lead time τ, the target is a binary spatial field

$$\mathbf{Y}_{t_0+\tau} \in \{0,1\}^{H\times W},$$

with $H = W = 512$, indicating convective-core occurrence on the native MSG grid over the western Sahel domain.

3.2.2 Model architecture for spatial core nowcasting

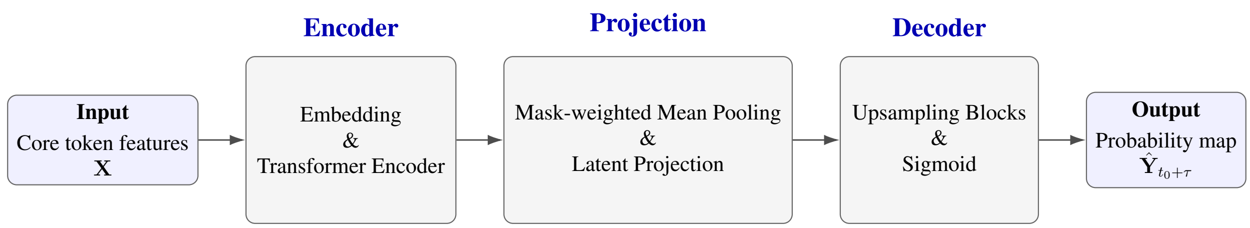


Figure 3: Schematic architecture of NCAST. The spatio-temporal input tensor $\mathbf{X}$ contains object-based representations of convective cores detected during the previous two hours. Core tokens are embedded and processed jointly by a spatio-temporal Transformer encoder. The resulting contextualised representations are aggregated using mask-weighted mean pooling to obtain a global latent representation, which is projected onto a spatial latent grid and decoded through convolutional upsampling blocks to produce a gridded probabilistic forecast of convective-core occurrence $\hat{\mathbf{Y}}_{t_0+\tau}$ at forecast lead time τ.

The spatio-temporal input tensor is reshaped into a set of $N_{\mathrm{tok}} = T \times N = 250$ tokens. Each token i is represented by a feature vector $\mathbf{x}_i \in \mathbb{R}^{13}$ and mapped into a latent space of dimension $d_{\mathrm{model}} = 128$ using a linear embedding layer

$$\mathbf{h}_i = \mathbf{W}_{\mathrm{in}}\mathbf{x}_i + \mathbf{b}_{\mathrm{in}},$$

followed by dropout with rate 0.2. Stacking these embeddings yields

$$\mathbf{H} \in \mathbb{R}^{N_{\mathrm{tok}}\times d_{\mathrm{model}}}.$$

Unlike the original Transformer formulation (Vaswani et al., 2017), no explicit positional encoding is applied. Spatial and temporal information are provided directly through core attributes (latitude, longitude, bounding-box coordinates, and lag-dependent temporal encodings). The encoder therefore operates in a permutation-invariant manner with respect to token ordering, as spatial and temporal information are encoded directly within the token attributes.

The embedded tokens are processed by a Transformer encoder with four layers and four attention heads per layer. Scaled dot-product attention within each head is defined as

$$\mathrm{Attn}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d_{\mathrm{head}}}} + \mathbf{M}\right)\mathbf{V},$$

where $\mathbf{Q} = \mathbf{H}\mathbf{W}_Q$, $\mathbf{K} = \mathbf{H}\mathbf{W}_K$, and $\mathbf{V} = \mathbf{H}\mathbf{W}_V$ are learned linear projections and $\mathbf{M}$ is a padding mask assigning $-\infty$ to logits corresponding to padded tokens.

The encoder outputs contextualised core representations

$$\tilde{\mathbf{H}} \in \mathbb{R}^{N_{\mathrm{tok}}\times d_{\mathrm{model}}}.$$

These are aggregated using mask-weighted mean pooling,

$$\mathbf{h}_{\text{pool}} = \frac{\sum_{i=1}^{N_{\text{tok}}} m_i \, \tilde{\mathbf{h}}_i}{\sum_{i=1}^{N_{\text{tok}}} m_i},$$

where $m_i \in \{0, 1\}$ indicates whether token i corresponds to a real core.

The pooled representation is first projected to a vector of length $128 \times 16 \times 16$ and subsequently reshaped into a spatial latent tensor

$$\mathbf{Z} \in \mathbb{R}^{128 \times 16 \times 16}.$$

This tensor is upsampled to the full 512×512 grid using a convolutional decoder composed of successive blocks of transpose convolutions, batch normalisation (Ioffe & Szegedy, 2015), ReLU activations (Nair & Hinton, 2010), and 2D dropout (Srivastava et al., 2014). A final 1×1 convolution produces logits

$$\mathbf{Z}^{\text{out}} \in \mathbb{R}^{H \times W},$$

which are transformed into probabilities via the sigmoid function

$$\hat{\mathbf{Y}}_{t_0+\tau} = \sigma(\mathbf{Z}^{\text{out}}).$$

Figure 3 provides a schematic illustration of the NCAST architecture. Training minimises a hybrid loss

$$\mathcal{L} = \alpha \, \mathcal{L}_{\text{BCE}} + (1 - \alpha) \, \mathcal{L}_{\text{spatial}},$$

with α and positive-class weight w_+, selected through hyperparameter optimisation using validation performance.

The weighted binary cross-entropy (BCE) term (Bishop, 2006) is

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} \left[w_+ \, y_{ij} \log \hat{y}_{ij} + (1 - y_{ij}) \log(1 - \hat{y}_{ij}) \right],$$

where y_{ij} and $\hat{y}_{ij}$ denote elements of $\mathbf{Y}_{t_0+\tau}$ and $\hat{\mathbf{Y}}_{t_0+\tau}$, respectively. In practice, this term is implemented using a numerically stable BCE-with-logits formulation.

The spatial consistency term, comparing locally averaged fields over a 9×9 neighbourhood (approximately 27 km on the native MSG grid), is aligned with established neighbourhood-based verification approaches such as the Fractions Skill Score (FSS; Roberts & Lean, 2008):

$$\bar{\hat{\mathbf{Y}}} = \text{AvgPool}_{9\times9}(\hat{\mathbf{Y}}), \qquad \bar{\mathbf{Y}} = \text{AvgPool}_{9\times9}(\mathbf{Y}),$$

$$\mathcal{L}_{\text{spatial}} = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} \left(\bar{\hat{y}}_{ij} - \bar{y}_{ij} \right)^2.$$

The number of Transformer layers, attention heads, and latent embedding dimension were fixed throughout the experiments to limit computational cost. Optimisation was performed using the AdamW optimiser (Loshchilov & Hutter, 2019).

A targeted one-factor-at-a-time hyperparameter search was performed independently for each forecast lead time around a reference configuration. The reference configuration used a learning rate of 10^{-4}, dropout probability of 0.2, positive-class weight $w_+ = 25$, and hybrid loss weighting parameter $\alpha = 0.3$. Each hyperparameter was then varied individually while keeping the others fixed, considering learning rates of 5×10^{-5}, 10^{-4}, and 2×10^{-4}; dropout probabilities of 0.1, 0.2, and 0.3; positive-class weights w_+ of 10, 25, and 50; and α values of 0.1, 0.3, and 0.5. Model selection and early stopping were based on validation AUC, using a patience of 5 validation evaluations and a minimum improvement threshold of 0.001. The best-performing NCAST configuration for each forecast lead time is summarised in Table 3.

Lead time	Learning rate	Dropout	Pos. weight	α	Val AUC
1 h	2×10^{-4}	0.2	25	0.3	0.976
3 h	1×10^{-4}	0.3	25	0.3	0.940
6 h	1×10^{-4}	0.3	25	0.3	0.914

Table 3: Best-performing hyperparameter configurations for each lead time based on validation AUC.

3.2.3 Ensemble forecasting

To reduce sensitivity to stochastic optimisation and random weight initialisation, a five-member ensemble was constructed for each forecast lead time. Here, ensemble refers to independently trained instances of the same model configuration and not to a meteorological ensemble generated from perturbations to initial conditions or model physics; member variability should therefore not be interpreted as meteorological forecast uncertainty. At inference time, the predicted probability fields $\hat{\mathbf{Y}}_{t_0+\tau}$ from the individual members were averaged pixel-wise to produce the final ensemble forecast. Unless otherwise stated, all NCAST results presented in this study correspond to the ensemble mean.

3.2.4 Ablation study

To assess the contribution of key components of the NCAST architecture, a series of ablation experiments were performed. Three aspects of the model were examined. First, the amount of temporal context was varied by considering inputs spanning one, two, three, or five time steps, corresponding to $\mathbf{X}_{t_0}$, $\{\mathbf{X}_{t_{-1}}, \mathbf{X}_{t_0}\}$, $\{\mathbf{X}_{t_{-2}}, \mathbf{X}_{t_{-1}}, \mathbf{X}_{t_0}\}$, and the full configuration $\mathbf{X}$, respectively. Second, the cyclical month and time-of-day encodings were removed to assess the contribution of temporal information. Third, the Transformer encoder was replaced by a feed-forward projection, thereby removing self-attention while leaving the decoder unchanged.

Each ablation variant was trained independently using the lead-time-specific hyperparameter configuration reported in Table 3. Hyperparameters were not re-optimised for individual ablation experiments, allowing the impact of each architectural modification to be assessed relative to the same reference configuration. To isolate the effect of each modification, ablation experiments were performed using a single NCAST model rather than the five-member ensemble. Model selection and early stopping followed the same procedure as the full model and were based on validation AUC with a patience of five validation evaluations. All variants were evaluated on the independent 2020–2024 test period.

4 Results

Convective-core occurrence prediction is a highly imbalanced binary classification problem, both for the single-location models and for the spatial NCAST framework (Appendix A). Performance is therefore assessed using complementary probabilistic metrics that quantify discrimination, calibration, overall accuracy, and spatial structure. These metrics are introduced here and applied consistently throughout the analyses that follow.

Discrimination is evaluated using the receiver operating characteristic (ROC) curve (Fawcett, 2006) and its area under the curve (AUC), which quantify the ability of the model to separate convective from non-convective cases across all probability thresholds by plotting the hit rate against the false-alarm rate. Higher AUC values indicate stronger discrimination. Calibration is examined using reliability diagrams, which compare forecast probabilities with observed event frequencies, revealing whether the model systematically overpredicts or underpredicts convective core occurrence across the forecast-probability range. Sharpness is assessed using the forecast-probability distributions shown alongside the reliability diagrams, indicating how frequently the model issues low-, medium-, and high-confidence forecasts.

Overall probabilistic accuracy is summarised by the Brier Score (Brier, 1950) and its skill score,

$$BS = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2, \qquad BSS = 1 - \frac{BS_{\text{model}}}{BS_{\text{ref}}}. \tag{1}$$

The BSS ranges from $-\infty$ to 1, with positive values indicating improvement over the reference forecast.

Because spatial displacement errors strongly affect convective verification, spatial structure is evaluated using the Fractions Skill Score (FSS) (Roberts & Lean, 2008). For a given neighbourhood size,

$$\text{FSS} = 1 - \frac{\sum_i \left(\bar{\hat{y}}_i - \bar{y}_i\right)^2}{\sum_i \left(\bar{\hat{y}}_i^2 + \bar{y}_i^2\right)},$$

where $\bar{\hat{y}}_i$ and $\bar{y}_i$ denote neighbourhood-averaged forecast and observed event fractions. The score ranges from 0 to 1, with higher values indicating better spatial correspondence.

No fixed probability threshold is applied in the primary evaluation. All metrics are computed using probabilistic forecasts directly, allowing model performance to be assessed across the full range of forecast probabilities rather than at a single decision threshold.

These metrics are applied as appropriate to each component of the study. For the single-location model, probabilistic evaluation is supplemented by AI explainability analyses, targeted sensitivity tests, and spatial transfer experiments. For the domain-wide NCAST model, pixel-wise BSS and neighbourhood-scale FSS are additionally used to assess forecast spatial organisation across the 1–6 hour lead-time range.

4.1 Single-location model performance

4.1.1 Predictive skill

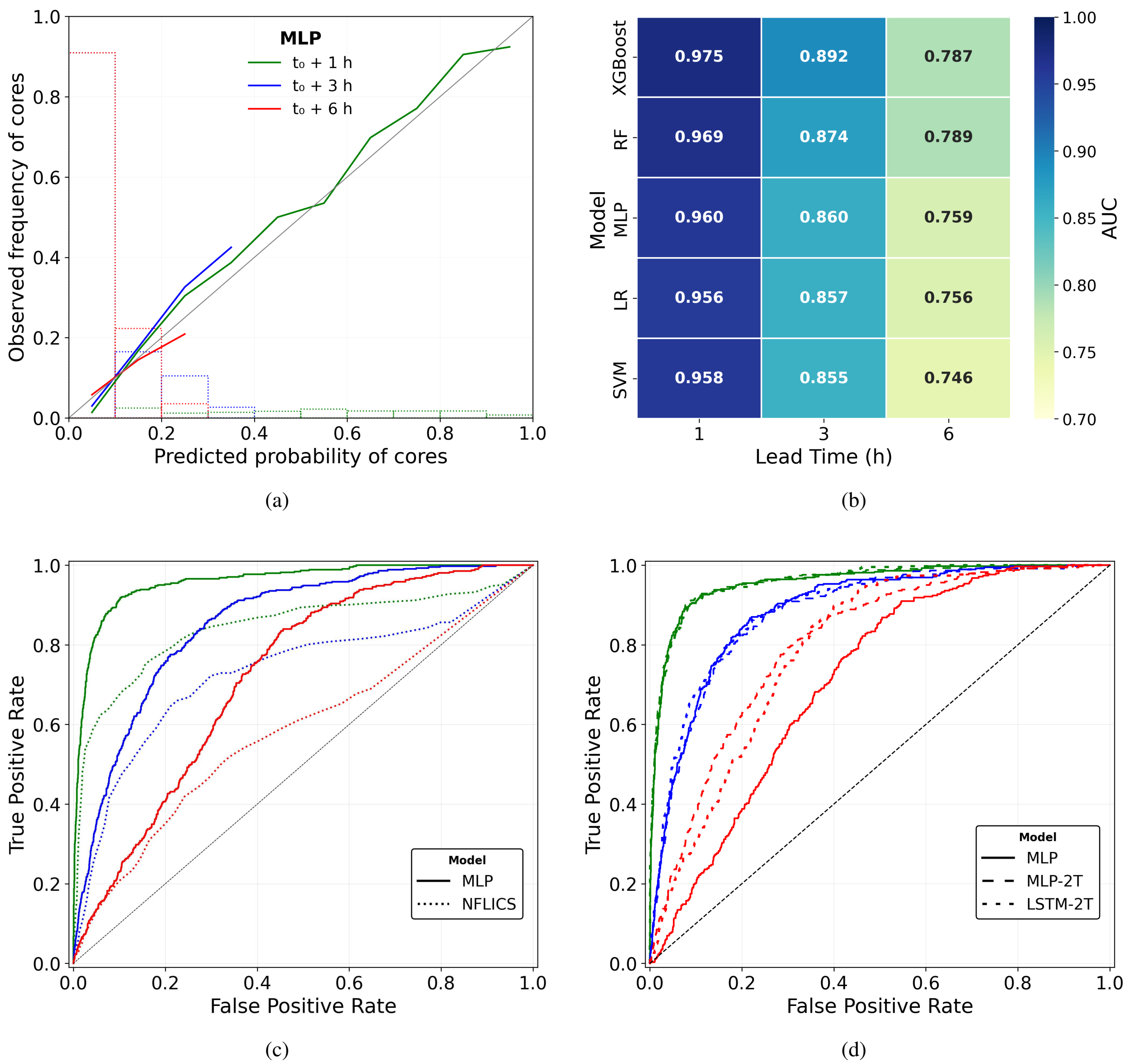


Figure 4: Performance of the single-location object-based framework over Dakar at forecast lead times of 1, 3, and 6 h (green, blue, and red, respectively). (a) Reliability and sharpness diagrams for the MLP; the grey diagonal indicates perfect reliability. (b) AUC values for the benchmark machine-learning models: Logistic Regression (LR), Support Vector Machine (SVM), Multilayer Perceptron (MLP), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). (c) ROC curves comparing the MLP with the NFLICS conditional climatology. (d) ROC curves comparing the MLP, MLP-2T, and LSTM-2T architectures. In panels (c) and (d), the diagonal line indicates no discrimination skill.

All benchmark models exhibit useful skill for local convective-core nowcasting (Figure 4b). XGBoost and Random Forest achieve the highest AUC values at all forecast lead times, while the MLP remains competitive despite its simpler architecture. These results demonstrate that the object-based core descriptors contain substantial predictive information for forecasts out to 6 hours ahead. The MLP is retained as the main model for the subsequent single-location analyses because the neural-network framework can be scaled up to explicitly model temporal evolution and interactions among populations of convective cores, as explored later using recurrent and attention-based architectures (Section 3.2.2).

The reliability diagrams (Figure 4a) indicate reasonable calibration at a 1-hour forecast lead time, with forecast probabilities broadly aligned with observed frequencies. At 3- and 6-hour forecast lead times, forecast distributions shift towards lower probabilities and exhibit reduced sharpness, consistent with declining predictability as forecast lead time increases.

Discrimination performance is shown in Figures 4c and 4d. The NFLICS reference forecast was evaluated using a neighbourhood scale of $L = 15$ km, matching the 5×5 target neighbourhood used to define convective-core occurrence for the single-location approach (Section 3.1.1). At this scale, the MLP substantially exceeds NFLICS at a 1-hour lead time. At 3 hours, discrimination remains strong, whereas the reference forecast degrades more rapidly. By 6 hours, the MLP retains useful skill while NFLICS approaches no-skill levels.

As shown in Figure 4d, incorporating information from t_{-1} improves discrimination at 3- and 6-hour lead times, indicating that even minimal temporal context enhances predictability. The comparable performance of the MLP-2T and LSTM-2T variants suggests that, for a two-time-step sequence, recurrent structure provides limited additional benefit beyond simple feature concatenation. The positive impact of temporal context motivates the use of longer temporal histories in the spatial NCAST framework, where multiple observations are used to represent short-term convective-core evolution.

4.1.2 Explainable AI: Shapley value analysis

To assess the contribution of each input feature to the predicted probability of convective-core occurrence over Dakar, SHAP (Shapley Additive Explanations) values (Lundberg & Lee, 2017) are used. These provide a theoretically grounded and internally consistent framework for interpreting complex machine-learning models. Rooted in cooperative game theory (Shapley, 1953), SHAP values quantify the marginal contribution of each feature by averaging its effect over all possible feature subsets.

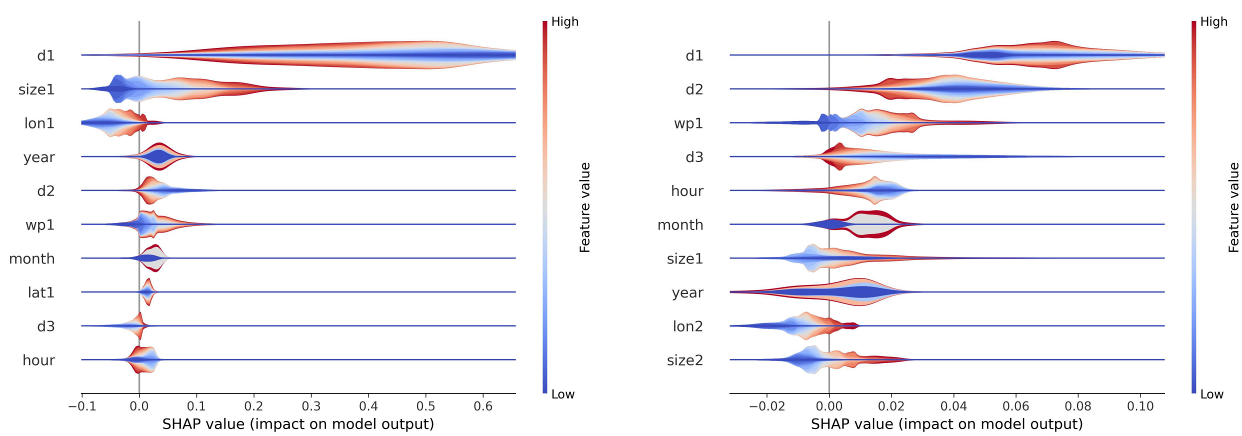
For the single-time-step MLP model, which uses only the core attributes available at the nowcast origin t_0, the input is given by the feature vector $\mathbf{X}_{t_0}$. The predicted probability $f(\mathbf{X}_{t_0})$ can be decomposed as

$$f(\mathbf{X}_{t_0}) = \mathbb{E}[f(\mathbf{X}_{t_0})] + \sum_{i=1}^{d} \phi_i, \tag{2}$$

where $\mathbb{E}[f(\mathbf{X}_{t_0})]$ denotes the expected model output over the chosen background distribution and ϕ_i is the Shapley value associated with feature i. Each ϕ_i is defined as

$$\phi_i = \sum_{S \subseteq \mathcal{I} \setminus \{i\}} \frac{|S|!\,(|\mathcal{I}| - |S| - 1)!}{|\mathcal{I}|!} \left[f(S \cup \{i\}) - f(S) \right], \tag{3}$$

where $\mathcal{I}$ denotes the index set of input features and S denotes any subset of $\mathcal{I}$ not containing feature i. SHAP values are computed on the test set to assess generalisation and to quantify feature importance in a manner consistent with operational use.



(a) (b)

Figure 5: SHAP summary plots for the ten most influential features in the single-time-step MLP model at forecast lead times of (a) 1 h and (b) 3 h. Features are ordered by mean absolute SHAP value, with colour indicating the feature magnitude.

We use the official SHAP implementation (Lundberg, 2018) with `DeepExplainer` to compute Shapley values for all lead times. `DeepExplainer` approximates exact Shapley values by propagating conditional expectations through the network and assumes independence between input features. Although this assumption is not strictly satisfied, given correlations among spatial and physical attributes (e.g. latitude, longitude, and distance), the resulting feature rankings remain stable and physically interpretable in this application. A random subset of the test set is used as the background distribution, providing a representative baseline while remaining computationally efficient compared with model-agnostic methods such as `KernelExplainer`.

Figure 5 shows SHAP summary plots for the ten most influential input features of the single-time-step MLP model at 1- and 3-hour lead times. At the 1-hour lead time, the distance to the nearest core (`d1`) dominates. Smaller distances (low `d1`, blue) are associated with the largest positive SHAP values, while larger distances contribute more weakly, indicating that short-range predictability is primarily controlled by proximity at t_0. This behaviour is consistent with persistence and short-term advection of nearby systems towards Dakar. Structural properties of the nearest core, particularly its area (`size1`) and wavelet power (`wp1`), also contribute positively, supporting the role of large and intense systems in sustaining convection over the site. The longitude of the nearest core (`lon1`) further reflects sensitivity to upstream positioning, with east–west location influencing whether convection is likely to approach or recede from Dakar.

At the 3-hour lead time, proximity remains the dominant factor, but its effect becomes less concentrated. The SHAP distribution for `d1` broadens and exhibits greater asymmetry, reflecting increased uncertainty in convective-core evolution at longer horizons. Distances to secondary cores (`d2` and `d3`) gain relative importance, indicating that the broader spatial configuration of convective activity becomes increasingly relevant. Structural intensity indicators (`wp1` and `size1`) continue to contribute positively, though with reduced magnitude, consistent with diminishing persistence effects beyond the immediate short range.

Temporal features gain relative importance at the 3-hour lead time. Both `hour` and `month` exhibit broader SHAP distributions, indicating stronger modulation by the diurnal cycle and seasonal background state. Since these variables are encoded as integers (`hour` = 0–23; `month` = 6–9), higher values of these variables are generally associated with positive SHAP contributions, corresponding primarily to afternoon and early evening hours and to the peak monsoon months (August–September). This pattern is consistent with the climatology of organised convection in the western Sahel.

To compare feature importance across lead times, the mean absolute SHAP value of each input feature is computed over the test set. Relative contributions are obtained by normalising the mean absolute SHAP values across physically interpretable features only (excluding padding masks, day-of-month variables, and the year variable). Features are grouped into seven categories—Distance, Size, Intensity, Season, Time of Day, Latitude, and Longitude—and category-level importance is calculated by summing normalised contributions within each group at each lead time. Figure 6 summarises the resulting relative importances.

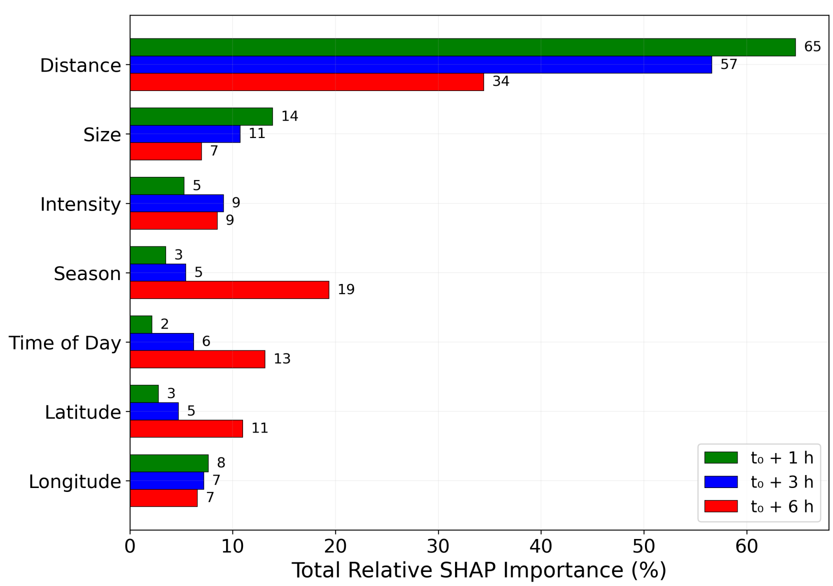


Figure 6: Relative SHAP importance of grouped features at 1-, 3-, and 6-hour forecast lead times. Each bar shows the normalised contribution of feature groups to model predictions, averaged over the test set.

At short lead times (1–3 h), predictability is governed primarily by core-specific attributes, with Distance accounting for more than half of the total relative importance. Size and Intensity provide additional contributions, indicating that the model also relies on information related to the persistence and organisation of nearby convective cores. By 6 h, the importance of Distance decreases substantially, while temporal and spatial context, particularly Season, Time of Day, and Latitude, become increasingly important. The increasing role of Latitude is consistent with the preferred meridional corridor of organised convective activity over the western Sahel (Laing & Fritsch, 1993; Taylor et al., 2017), which influences whether convective systems are favourably positioned to affect Dakar at longer lead times.

The SHAP analysis therefore reveals a coherent hierarchy of predictors: very short-range skill arises mainly from proximity and persistence, whereas longer-range skill reflects increasing modulation by background climatological and environmental conditions. While SHAP values quantify model reliance rather than causality, the alignment between learned feature importance and established convective dynamics supports the physical plausibility of the model behaviour.

4.1.3 Counterfactual spatial sensitivity analysis

To examine the model's sensitivity to the spatial placement of upstream convection, we construct controlled counterfactual inputs consisting of synthetic clusters of three convective cores. The three-core configuration matches the input structure of the single-time-step MLP and is intended to represent a compact organised convective system. The cluster is translated systematically across the context domain, with its centroid placed at each grid point surrounding Dakar. The two additional cores are generated using small Gaussian perturbations (standard deviation 0.2°) with a fixed random seed to preserve spatial coherence. Wavelet power and core area are fixed to their median values in the training set. All remaining non-spatial attributes are held constant except in experiments that explicitly examine seasonal or diurnal effects. Consequently, variations in the predicted probability arise solely from controlled changes in the spatial configuration of the synthetic cluster.

At each grid point, the model-predicted probability of convective-core occurrence over Dakar is evaluated conditional on placing the cluster centroid at that location at the nowcast origin time t_0. Sensitivity is defined as the difference between this conditional probability and the corresponding NFLICS unconditional climatological probability for Dakar at the same lead time and target hour, evaluated using the same 15 km spatial scale adopted in the target definition.

For visualisation, probability differences are smoothed using a Gaussian filter ($\sigma = 1$) to highlight coherent spatial patterns. Positive values indicate locations where upstream convection increases the predicted likelihood of convective-core occurrence over Dakar relative to climatology, whereas negative values indicate locations associated with reduced likelihood.

Figure 7 presents sensitivity fields for three representative nowcast-origin times on 15 July 2020 (1200, 1500, and 2100 UTC) across 1-, 3-, and 6-hour lead times. Figure 8 shows the corresponding seasonal sensitivity patterns for June–September at a fixed origin time (1500 UTC). At short lead times, the sensitivity patterns are qualitatively similar in both figures. At the 1-hour lead time, positive sensitivity is highly localised around Dakar and confined to a narrow region immediately upstream of the city, consistent with a near-deterministic regime dominated by short-range advection and persistence. Weakly negative sensitivities at larger distances indicate configurations for which upstream convection is unlikely to influence Dakar within one hour and therefore contributes little additional predictive information beyond climatology.

By 3 hours, the region of positive sensitivity broadens modestly inland while remaining centred near Dakar. Spatial gradients weaken and become more heterogeneous, indicating increased uncertainty in convective-core evolution and propagation. Sensitivity is enhanced to the east and south-east of Dakar, suggesting that prior convection in these sectors is more likely to result in a core over Dakar three hours later. This asymmetry is consistent with the climatological westward movement of organised convection associated with the African Easterly Jet.

At the 6-hour lead time, the sensitivity patterns become qualitatively different. Sensitivity patterns become substantially broader, with enhanced values extending far inland across southern Mauritania and northern Mali. This inland expansion and east–west asymmetry are consistent with the preferred westward propagation and organisation of Sahelian convective systems (Laing & Fritsch, 2000; Mathon et al., 2002; Maranan et al., 2018). Spatial gradients weaken further, indicating reduced deterministic influence of individual upstream features at this horizon.

At this lead time, remote cores to the east are associated with increased probability over Dakar, whereas cores located close to Dakar six hours earlier tend to suppress subsequent development. This behaviour is consistent with documented recharge–discharge characteristics of Sahelian convection, whereby deep convection locally consumes available instability and reduces the likelihood of renewed development until instability has recovered (Taylor et al., 2017; Birch et al., 2013). Concentric temporal organisation of rainfall systems of this type has also been studied by Bassford et al. (2025).

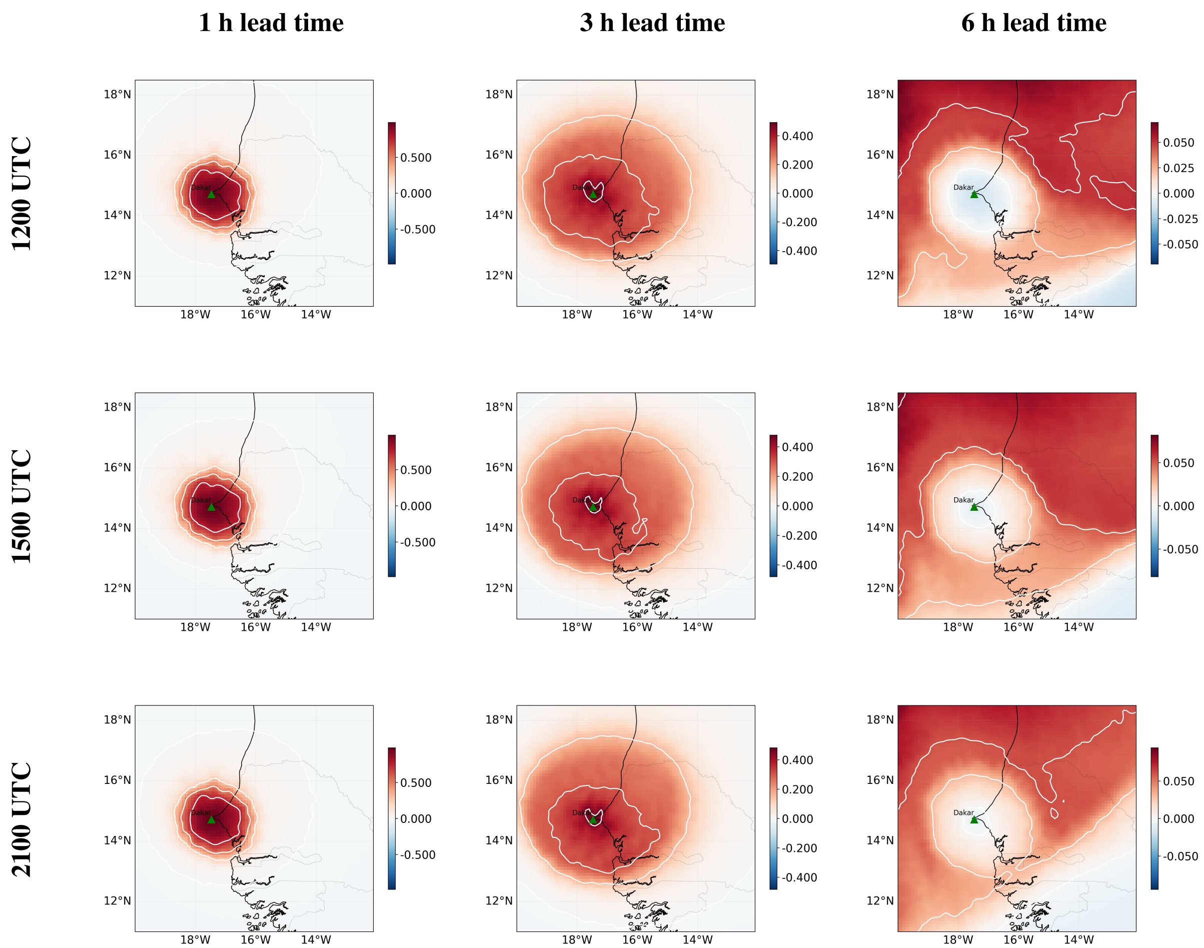


Figure 7: Spatial sensitivity of the single-location MLP model to the placement of a synthetic three-core cluster on 15 July 2020. Panels show probability differences relative to NFLICS unconditional climatology. Columns correspond to lead times of 1, 3, and 6 h; rows to nowcast origin times (1200, 1500, and 2100 UTC). Colour scales differ across lead times.

Diurnal modulation is most evident at the 6-hour lead time (Figure 7). Sensitivity fields show an advancing region of negative probability difference progressing from the south-east between $t_0 = 1200$ and 2100 UTC. The boundary between enhanced probability to the north and west and reduced probability to the south and east is sharp and approximately linear, propagating roughly 450 km over 9 hours (approximately 14 $\mathrm{m\,s^{-1}}$). This behaviour is consistent with the influence of convectively generated gravity waves and mesoscale cold-pool outflows on the organisation and timing of deep convection over West Africa (Birch et al., 2013).

A clear seasonal modulation is also evident at the 6-hour lead time (Figure 8). During August, positive sensitivity regions are more extensive and coherent, extending farther inland and showing stronger values to the south. This pattern is consistent with enhanced low-level moisture, deeper instability, and a more intense and zonally coherent African Easterly Jet during the core monsoon period (Laing & Fritsch, 2000; Maranan et al., 2018), conditions that favour sustained westward propagation of organised convection. In contrast, June and July exhibit weaker and more spatially confined sensitivity patterns, reflecting the transitional structure of early-monsoon convection and reduced mesoscale organisation (Mathon et al., 2002). In these earlier monsoon stages, a core to the north-east indicates locally enhanced moisture and favourable background conditions, increasing the likelihood of subsequent development near Dakar. September displays renewed sensitivity to the south and east, but also positive sensitivity over Dakar itself, indicating reduced local suppression compared with earlier months as large-scale monsoon circulation begins to weaken.

These proposed mechanisms should be regarded as physically motivated hypotheses rather than definitive dynamical attribution. A more formal conceptual framework for interpreting the conditional probability fields is presented in Section 5.1.

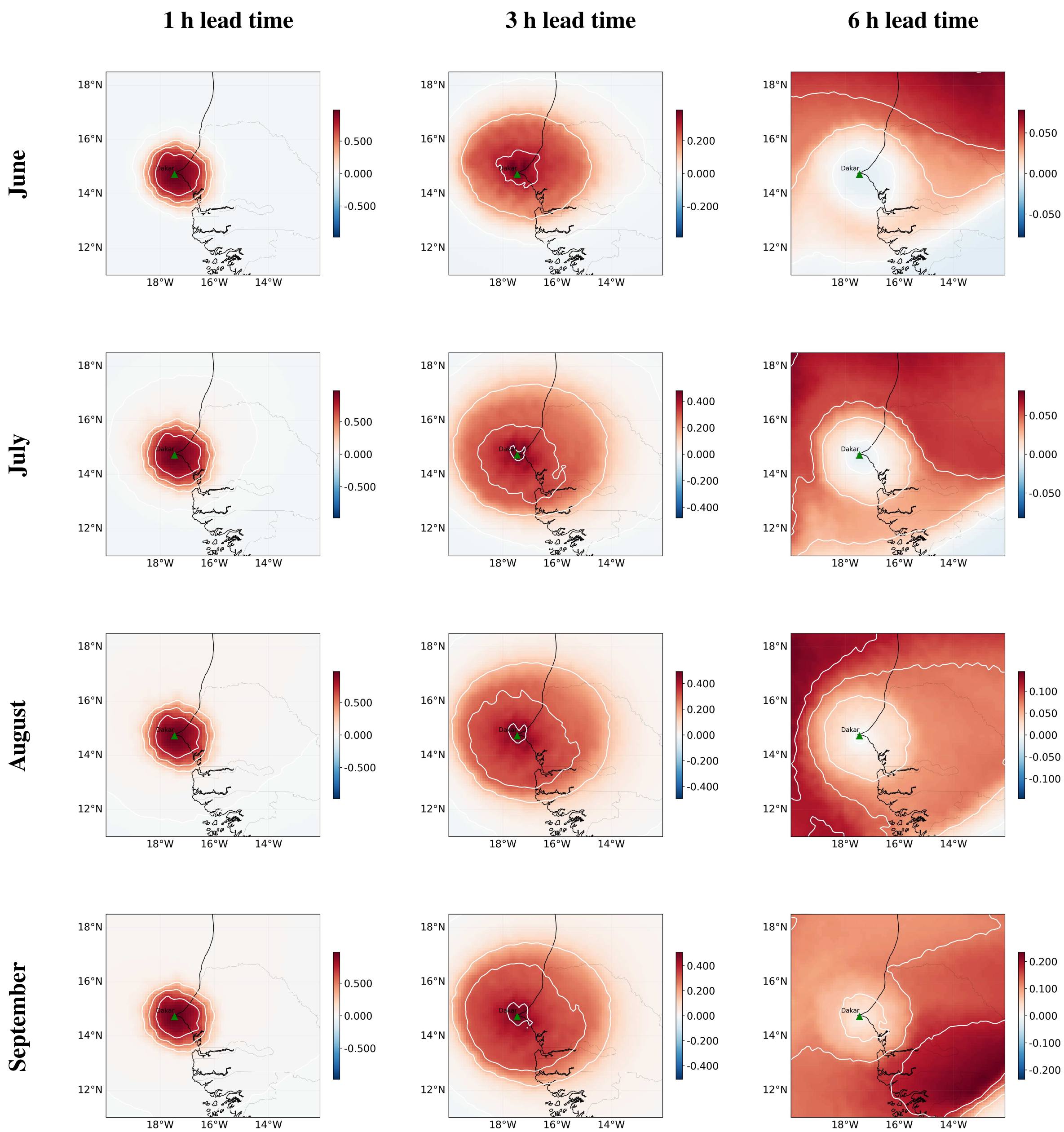


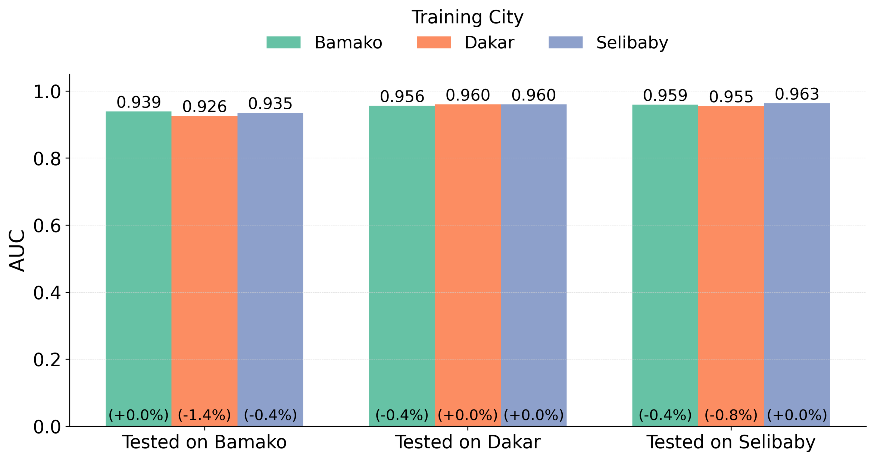
Figure 8: Seasonal variation in spatial sensitivity of the MLP nowcasting model to synthetic upstream convection, shown as probability differences relative to NFLICS unconditional climatology. Columns denote lead times (1, 3, 6 h); rows denote months (June–September). The nowcast origin time (1500 UTC) and year (2020) are fixed. Colour scales differ across lead times.

These counterfactual sensitivity experiments provide a spatial interpretation of the feature-importance hierarchy identified earlier. Strong localisation at short lead times aligns with the dominance of proximity-based information, whereas the inland expansion of sensitivity at longer horizons reflects increasing modulation by environmental state and convective organisation. The patterns represent conditional sensitivities learned by the model rather than explicit dynamical causation; however, their consistency with established convective dynamics suggests that the model has learned physically meaningful spatial relationships. This type of controlled spatial perturbation analysis is a direct advantage of the object-based formulation, allowing individual convective core to be perturbed and interpreted explicitly—an approach substantially more difficult with purely image-based nowcasting models.

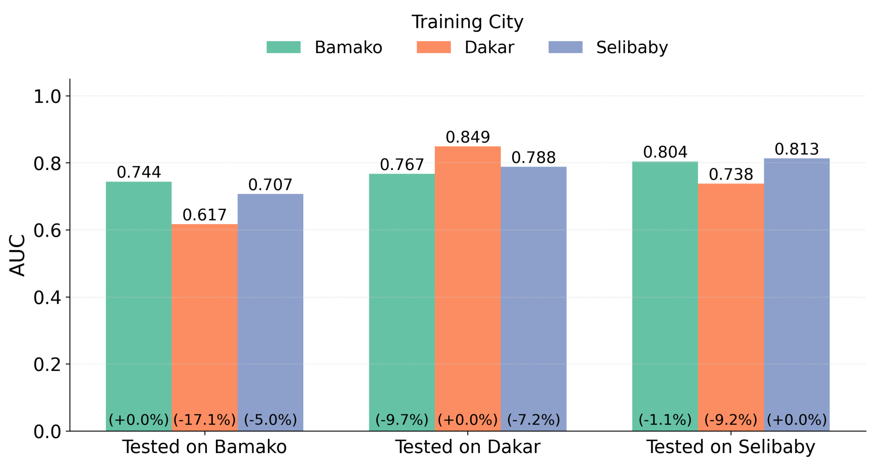
4.1.4 Single-location training and spatial transferability

To evaluate spatial transferability, we trained single-time-step MLP models using data from one location and evaluated them both locally and at other locations without retraining. The model architecture, preprocessing, optimisation settings, and lead-time-specific hyperparameters are identical to those of the best-performing MLP configuration described in Section 3.1. No structural modifications are introduced for this experiment.

To enable cross-location application, only the coordinate representation is altered: the latitude and longitude of each convective core are expressed as offsets relative to the target location, i.e. $\Delta\mathtt{lat}_k = \mathtt{lat}_k - \mathtt{lat}_{\text{site}}$ and $\Delta\mathtt{lon}_k = \mathtt{lon}_k - \mathtt{lon}_{\text{site}}$ for each core index k. All other features, including distance, size, wavelet power, and timestamp components at t_0, remain unchanged. Core attributes for all sites are extracted using the same satellite observations and detection procedure. Apart from the training location itself, the modelling pipeline is identical across experiments. This design isolates whether the model captures transferable atmospheric relationships rather than memorising site-specific feature distributions.



(a)



(b)

Figure 9: Cross-location generalisation of the single-time-step MLP model at (a) 1-hour and (b) 3-hour lead times. Bars show AUC when training on one city (colour) and testing on another. Values in parentheses denote the change in AUC relative to training and testing on the same city (ΔAUC). The model architecture is identical in all cases; only the training location differs.

Figure 9 shows strong cross-location generalisation at short lead times. At 1 hour, all models achieve AUC values exceeding 0.92 regardless of training location, indicating only minor transfer penalties. This indicates that the relative-coordinate formulation captures broadly transferable aspects of short-range convective behaviour. At 3 hours, discrimination decreases and transfer becomes more challenging, but models trained on Bamako and Selibaby retain useful performance when applied to other sites (AUC > 0.71). In contrast, the model trained on Dakar exhibits the largest degradation when evaluated inland, suggesting that coastal–inland differences in convective organisation and background forcing may reduce portability at longer horizons. These results indicate that the object-based representation encodes predictive information that generalises across locations at short lead times, while regional dependencies become increasingly influential as forecast lead time increases.

4.2 Evaluating NCAST for the western Sahel

The following analyses evaluate NCAST over the independent 2020–2024 test period using all available scenes during the June–September monsoon season. For comparisons involving NFLICS, the ground-truth and persistence fields are evaluated using a neighbourhood-based definition of convective occurrence with neighbourhood size $L = 75$ km, consistent with the operational implementation of the conditional-climatology benchmark.

4.2.1 Ensemble performance

To assess the effect of random initialisation and ensemble averaging, the performance of the five individual members was compared with that of the ensemble forecast. Figure 10 shows that variability between ensemble members remains small, although the spread increases slightly at the 6-hour lead time. The ensemble forecast consistently improves AUC relative to each individual member, indicating enhanced discrimination and reduced sensitivity to random initialisation. Improvements in FSS are generally smaller and less consistent across lead times. The ensemble forecast is therefore used for all subsequent evaluations.

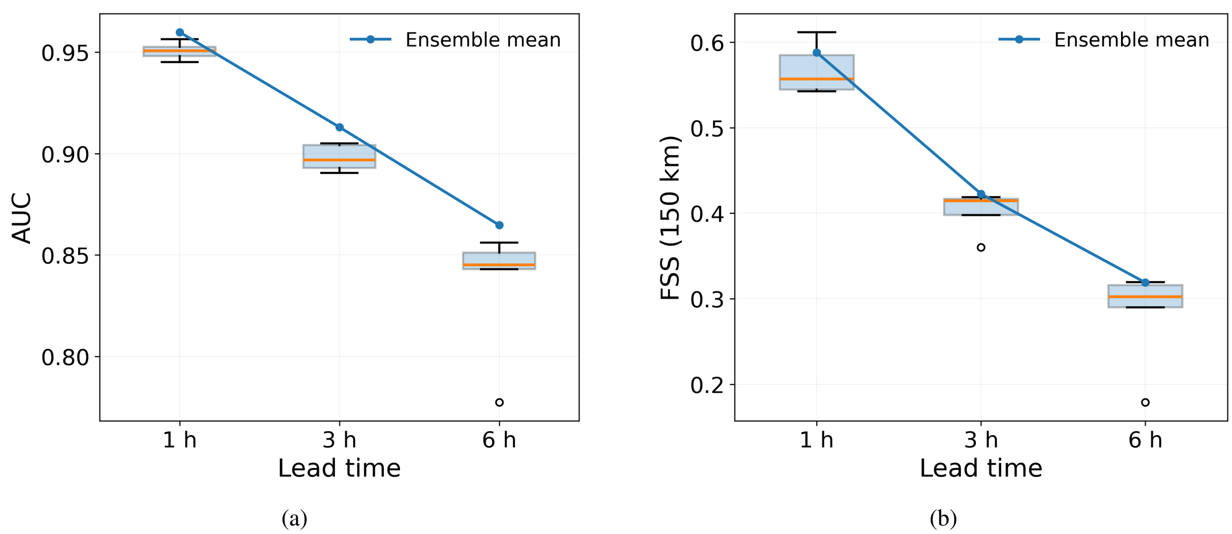


Figure 10: Performance of the five-member NCAST ensemble at lead times of 1, 3 and 6 h. Boxplots show the distribution of scores across the five independently trained ensemble members, while the line shows the ensemble forecast obtained by pixel-wise averaging of member probabilities prior to evaluation. Ensemble averaging consistently improves AUC, while FSS remains comparable to the distribution of individual members. (a) Area under the receiver operating characteristic curve (AUC). (b) Fractions Skill Score (FSS) evaluated at a spatial scale of 150 km.

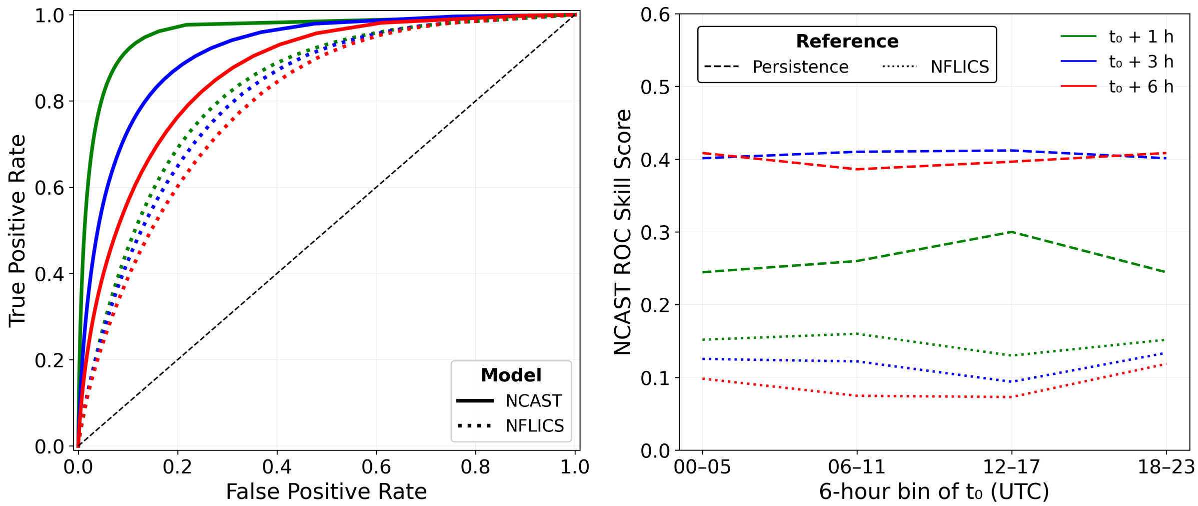
4.2.2 Nowcast skill relative to benchmarks

Figure 11 summarises the pixel-wise discrimination skill of NCAST over the western Sahel. Panel (a) presents ROC curves for 1-, 3-, and 6-hour forecasts, with AUC values of 0.96, 0.91, and 0.89, respectively. Corresponding NFLICS AUC values are approximately 0.82, 0.80, and 0.78, indicating substantially weaker discrimination at all lead times. Although discrimination decreases gradually with forecast horizon for both models, NCAST maintains substantially stronger separation between convective and non-convective pixels out to 6 hours.

Panel (b) shows the ROC-based skill score (ROCSS) of NCAST relative to two reference forecasts: persistence and the operational NFLICS conditional climatology, computed in 6-hour bins of nowcast origin. ROCSS is defined as

$$\mathrm{ROCSS} = 1 - \frac{\mathrm{AUC}_{\mathrm{ref}}}{\mathrm{AUC}_{\mathrm{NCAST}}},$$

so larger values correspond to greater added discrimination skill from NCAST, while smaller values occur when the reference forecast already achieves comparatively high discrimination. This formulation measures fractional improvement in AUC relative to the reference forecast.



(a) (b)

Figure 11: Pixel-wise evaluation of NCAST over the western Sahel. Forecast lead times of 1, 3, and 6 h are shown in green, blue, and red, respectively. Panel (a) shows ROC curves for NCAST (solid) and NFLICS (dotted). Panel (b) shows ROCSS relative to persistence (dashed) and the operational NFLICS conditional climatology (dotted), stratified by 6-hour bins of nowcast origin.

ROCSS is positive for all lead times and all time-of-day windows, indicating consistent added discrimination relative to both persistence and NFLICS throughout the diurnal cycle. The largest gains relative to persistence occur at 3–6 hours, reflecting the rapid degradation of short-term persistence at these horizons, whereas NCAST maintains comparatively strong separation skill. At the 1-hour lead time, gains relative to persistence are most pronounced during 12–17 UTC, when rapid afternoon convective development reduces persistence performance. Relative to NFLICS, added skill is smaller in magnitude but remains positive across all lead times. Diurnal variation in ROCSS is modest, with slight reductions during 12–17 UTC when the location-conditioned climatology performs relatively well. Outside this window, improvements are slightly larger but broadly consistent across time-of-day bins.

Spatial forecast quality is assessed using the Fractions Skill Score (FSS), evaluated across neighbourhood scales ranging from approximately 10 km (3 × 3 MSG grid cells) to 250 km (81 × 81 MSG grid cells) and forecast lead times. Scores are averaged over all verification instances and are shown in Figure 12.

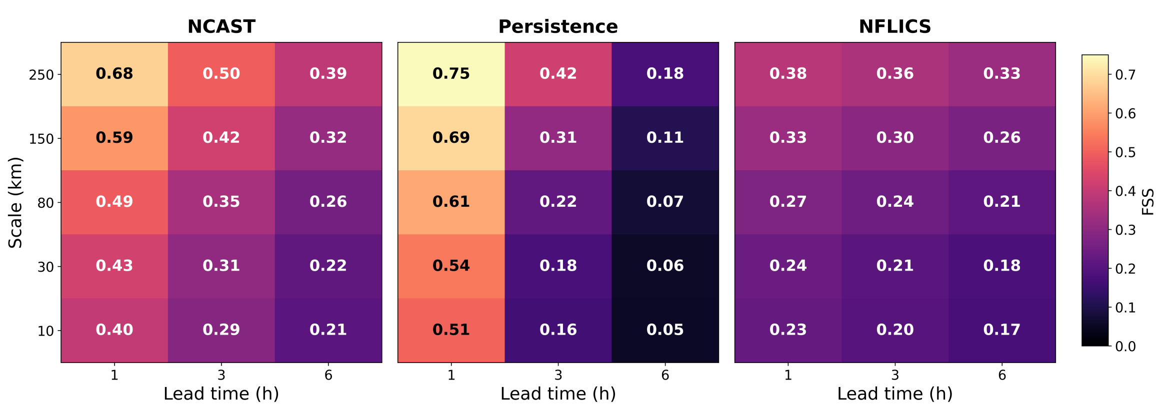


Figure 12: Fractions Skill Score for NCAST, persistence, and NFLICS as a function of spatial scale and lead time.

NCAST maintains useful spatial skill across a wide range of neighbourhood scales and forecast lead times (Figure 12). FSS increases with neighbourhood scale for all lead times, approaching 0.5 at scales of approximately 80 km and larger for the 1-hour forecast. These scales are representative of mesoscale and organised convective systems. Although FSS decreases with lead time, the reduction is gradual, indicating that NCAST preserves coherent spatial organisation of convective systems at longer lead times.

Persistence exhibits relatively high FSS at the 1-hour lead time, particularly at larger neighbourhood scales, reflecting strong short-range spatial overlap between forecasted and observed convective cores. However, FSS declines rapidly with lead time, indicating limited ability to maintain spatial correspondence beyond the immediate forecast range. NFLICS exhibits comparatively stable but lower FSS values, with weak sensitivity to both lead time and spatial scale, consistent with its climatological, location-conditioned formulation.

To assess probabilistic forecast skill across the domain, Figure 13 shows the pixel-wise Brier Skill Score (BSS) of NCAST relative to NFLICS, stratified by forecast lead time and by 6-hour bins of the nowcast origin time t_0.

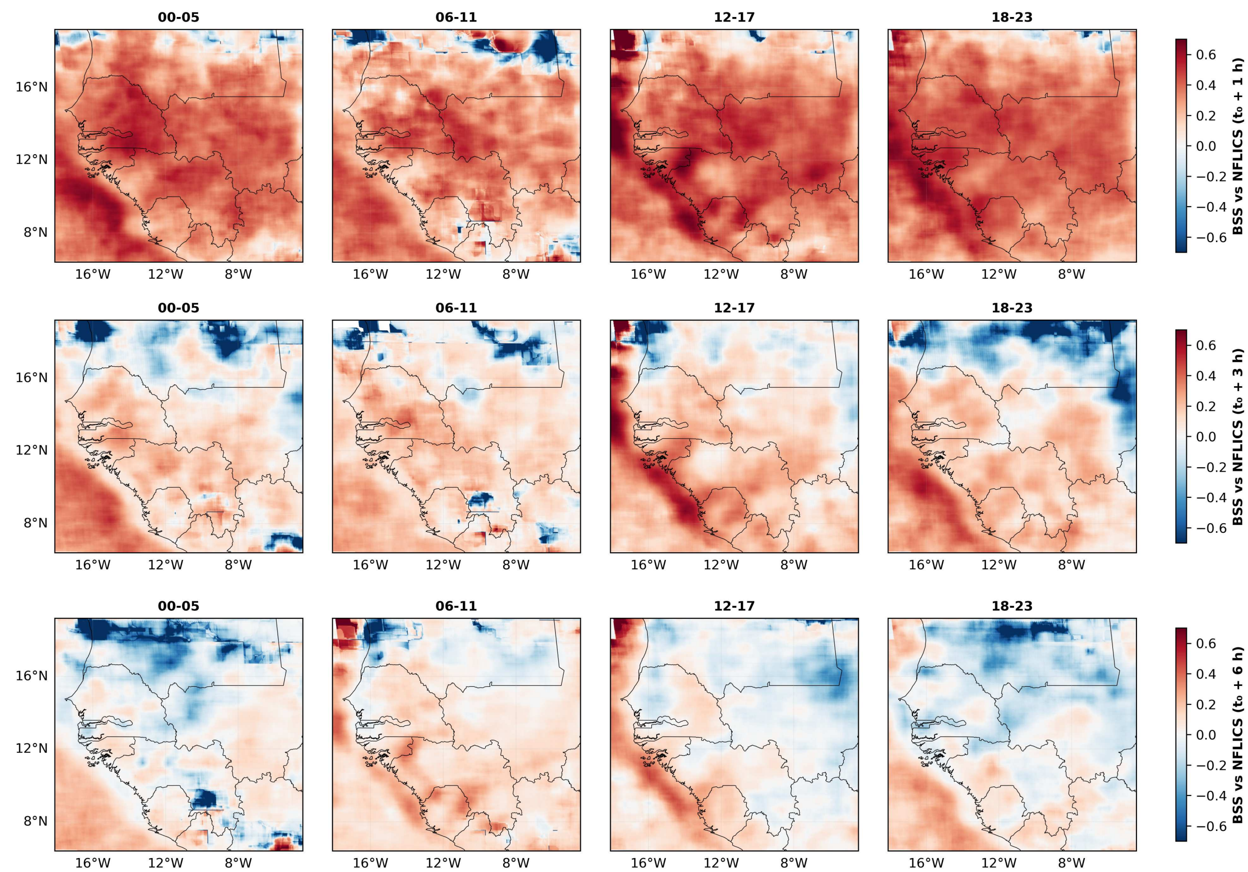


Figure 13: Spatial distribution of the pixel-wise Brier Skill Score (BSS) of NCAST relative to NFLICS over the western Sahel, shown for lead times of 1, 3, and 6 h and averaged by 6-hour bins of the nowcast origin (UTC). Rows correspond to forecast lead times and columns to 6-hour bins of the nowcast origin. Warm colours indicate positive BSS (added skill of NCAST relative to NFLICS), cool colours indicate negative BSS, and white corresponds to near-zero skill difference.

At the 1-hour lead time, positive BSS values are widespread across the convectively active belt of the western Sahel, with particularly strong gains along the west coast. At 3 hours, positive BSS remains concentrated within the primary organised-convection corridor, while by 6 hours it becomes more spatially confined, consistent with increasing forecast uncertainty at longer horizons. Coherent regions of positive BSS nevertheless persist within the organised convective belt at all lead times (Laing & Fritsch, 1993; Mathon et al., 2002; Parker et al., 2005).

Lower or locally negative BSS values occur more frequently north of approximately 16°N, where deep convection is comparatively infrequent (Nicholson, 2009). In these low base-rate environments, small forecast errors and false alarms more readily reduce apparent skill relative to the conditional-climatology reference. This indicates that performance degradation is regime-dependent rather than spatially uniform across the domain.

Across all lead times, NCAST exhibits particularly strong positive BSS for forecasts verifying during the 12–17 UTC window, corresponding to the peak phase of convective development. Because the analysis is stratified by t_0, this regime corresponds to initialisation times of 11–16 UTC (1 h), 09–14 UTC (3 h), and 06–11 UTC (6 h). NCAST therefore retains added probabilistic skill relative to NFLICS when forecasts verify during the most dynamically active phase of the diurnal cycle.

4.2.3 Example NCAST forecast

Figure 14 presents an example NCAST forecast for a multi-core convective system over the western Sahel.

The top row displays observed cores at $t_0 - 2$ h, $t_0 - 1$ h, and t_0. For clarity, only hourly frames are shown, although the model ingests core tokens at 30-minute resolution. The sequence highlights westward propagation and increasing spatial organisation prior to forecast initialisation.

The bottom row shows the corresponding forecasts at $t_0 + 1$ h, $t_0 + 3$ h, and $t_0 + 6$ h. At the 1-hour lead time, high probabilities remain concentrated near the observed convective cores. By 3 hours, the forecast becomes more spatially extensive while retaining the dominant organisation of the convective system. At 6 hours, the forecast becomes increasingly diffuse, reflecting greater uncertainty in the location and structure of the convection.

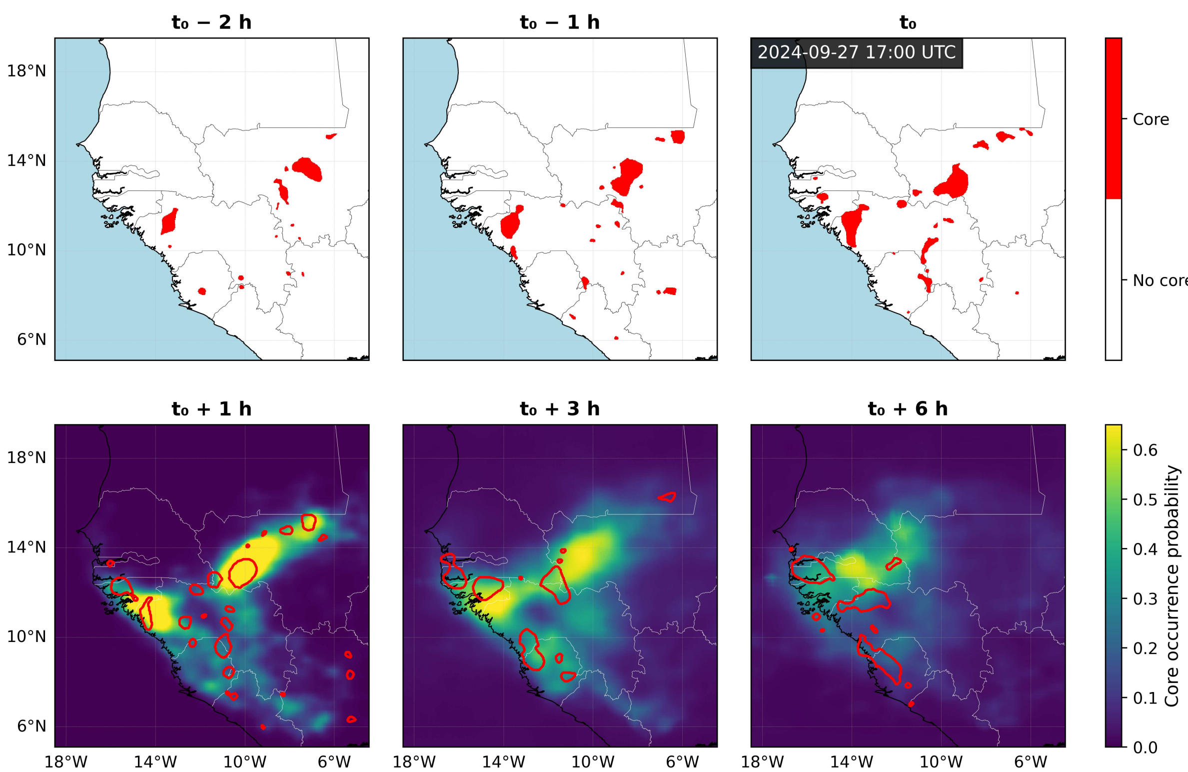


Figure 14: Example NCAST forecast for a multi-core convective system over the western Sahel, initialised at $t_0 = 27$ September 2024, 17:00 UTC. Top row: observed convective cores (filled red) at $t_0 - 2$ h, $t_0 - 1$ h, and t_0. Bottom row: probabilistic forecasts valid at verification times $t_0 + 1$ h, $t_0 + 3$ h, and $t_0 + 6$ h. Observed convective cores at the corresponding verification times are shown as unfilled red outlines.

Exact pixel-level reproduction of individual cores is not expected due to the object-based representation and probabilistic formulation. Nevertheless, the forecast retains coherent mesoscale structure and captures the dominant propagation and organisation of the convective system across the 1–6 h forecast range. This example illustrates how NCAST translates object-based core information into spatially distributed probabilistic forecasts of convective evolution.

4.2.4 NCAST ablation study

Figure 15 summarises the impact of key architectural components on forecast skill. Removing self-attention results in the largest degradation in AUC and one of the largest degradations in FSS across all lead times, demonstrating that modelling interactions between convective cores is important for accurate nowcasting. The reduction is particularly pronounced at $t_0 + 1$ h, where the AUC decreases from approximately 0.95 to 0.93 and the FSS decreases from approximately 0.58 to 0.48.

The contribution of temporal features increases with lead time. At the 1-hour lead time, removing the temporal encodings has only a small impact on performance, whereas at the 6-hour lead time the AUC decreases from approximately 0.86 to 0.81. This suggests that seasonal and diurnal information becomes increasingly important as the influence of the current convective-core configuration weakens.

The comparison of temporal input configurations shows that the model remains skilful when only the current observation $\mathbf{X}_{t_0}$ is provided. Adding one previous time step $\{\mathbf{X}_{t_{-1}}, \mathbf{X}_{t_0}\}$ yields a small improvement in both AUC and FSS, particularly at 1 and 3 h lead times. Incorporating an additional time step $\{\mathbf{X}_{t_{-2}}, \mathbf{X}_{t_{-1}}, \mathbf{X}_{t_0}\}$ produces little further benefit and, in some cases, slightly reduces performance. The differences between the reduced-history configurations and the full NCAST model are generally modest, suggesting that much of the predictive information is contained within the most recent convective-core configuration. The limited gains from longer temporal histories may reflect the use of average temporal pooling and the absence of explicit temporal positional encodings.

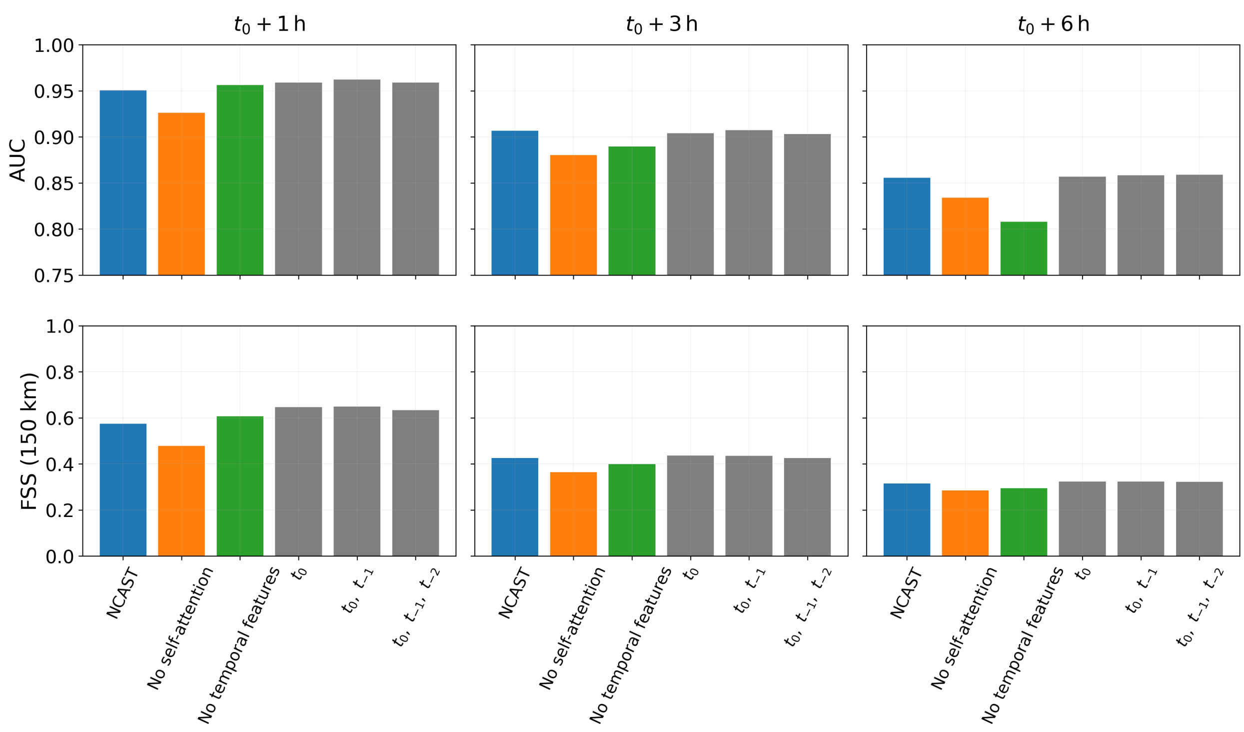


Figure 15: Ablation study of NCAST at forecast lead times of 1, 3, and 6 h. The top row shows the area under the receiver operating characteristic curve (AUC) and the bottom row shows the Fractions Skill Score (FSS) evaluated at a spatial scale of 150 km. The full model is compared with variants without self-attention, without temporal features, and with reduced temporal input histories. Removing self-attention produces the largest reduction in both discrimination and spatial skill, highlighting the importance of modelling interactions between convective cores. Temporal features become increasingly important at longer lead times. In contrast, extending the temporal input history beyond one previous observation yields only modest additional improvements with the current architecture.

5 Discussion

5.1 Interpretation of the conditional probability maps

The patterns shown in Figure 7 exhibit coherent spatial structure that aligns with known aspects of West African convective behaviour. These structures illustrate how a data-driven conditional climatology can encode physically interpretable dynamics. To interpret these patterns, consider the conditional probability field

$$P_c(\mathbf{s}, t_1 \mid \mathbf{s}_0, t_0),$$

which denotes the probability of a convective core at location $\mathbf{s}$ and time t_1, given that the closest prior core at time t_0 was located at $\mathbf{s}_0$. In this framework, Figure 7 displays $P_c(\mathbf{s}, t_1 \mid \mathbf{s}_0, t_0)$ as a function of the prior-core location $\mathbf{s}_0$, with $(\mathbf{s}, t_1, t_0)$ fixed.

In an idealised isotropic environment, for example in the absence of ambient wind, one expects radial symmetry such that

$$P_c = F(r, \Delta t),$$

where $r = \|\mathbf{s} - \mathbf{s}_0\|$ and $\Delta t = t_1 - t_0$. Under this assumption, the conditional probability depends only on separation distance, and the geometry is symmetric in $\mathbf{s}$ and $\mathbf{s}_0$. The resulting structure consists of concentric level sets in prior-core space, as illustrated schematically in Figure 16(a). Physically, a localised convective event may generate cold pools and gravity waves that spread radially, potentially triggering convection at remote locations while suppressing it nearby. A simple representation of such behaviour is an expanding ring,

$$P_c = P_0 \exp\left(-\frac{(r - c\,\Delta t)^2}{2\sigma^2}\right),$$

where c denotes a characteristic propagation speed and σ controls radial spread. The evolution in Figure 7 is qualitatively consistent with this behaviour: strongly localised at short lead times and expanding outward as Δt increases.

Beyond radial structure, Figure 7 also reveals a linear region of reduced conditional probability that shifts systematically with Δt. Observational and modelling studies document organised convective structures and associated wave-like disturbances in West Africa, often linked to the African Easterly Jet and mesoscale organisation (Birch et al., 2013; Parker et al., 2005). A simplified anisotropic representation is

$$P_c = G[(\mathbf{s}_0 + \mathbf{c}\,\Delta t - \mathbf{s}) \cdot \mathbf{n}] = G[c\,\Delta t - (\mathbf{s} - \mathbf{s}_0) \cdot \mathbf{n}],$$

where $\mathbf{c}$ is a propagation vector and $\mathbf{n}$ a unit vector normal to the disturbance. In this case, level sets of P_c are approximately linear in prior-core space, as illustrated in Figure 16(b), reflecting directional propagation and anisotropy introduced by environmental flow. The displacement of these features between successive lead times is qualitatively consistent with propagation speeds reported for mesoscale disturbances in the region.

Although the sensitivity experiments employ a synthetic three-core cluster to represent a compact organised system, the interpretation here treats $\mathbf{s}_0$ as the centroid of that cluster. This simplification allows the conditional probability structure to be analysed in terms of an effective prior-core location without altering the spatial characteristics of the patterns discussed.

The observed conditional probability fields can be interpreted as the superposition of an expanding concentric structure and a propagating anisotropic disturbance. Both mechanisms are well documented in studies of West African convection. Their emergence here from a purely data-driven conditional framework suggests that the resulting probability fields encode physically meaningful organisation and propagation characteristics rather than arbitrary statistical artefacts. This interpretation is offered as a physically grounded heuristic rather than a dynamical derivation, but the consistency between learned structure and established convective behaviour supports the physical plausibility of the framework.

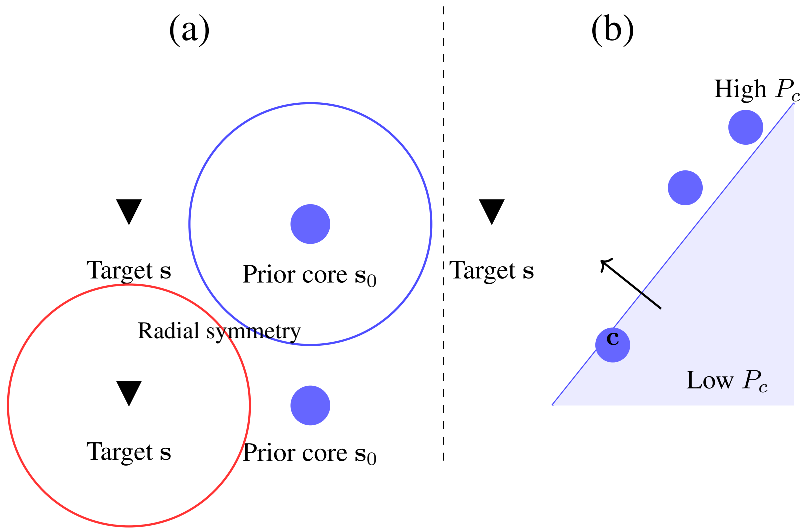


Figure 16: Schematic interpretation of the conditional probability field $P_c(\mathbf{s}, t_1 \mid \mathbf{s}_0, t_0)$. (a) In an isotropic environment, the conditional probability depends only on radial distance $r = \|\mathbf{s} - \mathbf{s}_0\|$ and is maximised along an expanding ring centred on the prior core. (b) With propagation in direction $\mathbf{c}$, anisotropy arises: the conditional probability is structured by the projection $(\mathbf{s} - \mathbf{s}_0) \cdot \mathbf{n}$, producing linear level sets in prior-core space.

5.2 Computational requirements, operationalisation, and limitations

The computational cost of the framework is dominated by the convective-core identification stage. In the operational nowcasting system, the wavelet-based core-detection algorithm requires approximately 80 s to process a full MSG disk, including ancillary processing tasks such as handling additional satellite fields and data archiving. The operational configuration uses six wavelet scales rather than the 45 scales employed by Klein et al. (2018) for scale-sensitivity analyses, substantially reducing computational cost. Including image acquisition and processing overheads, end-to-end core detection is typically completed within two minutes of image availability.

The single-location formulation is computationally lightweight and well suited to short forecast lead times. Both training and inference can be performed on standard consumer-grade hardware without a GPU, with training completing in minutes on a typical CPU. Owing to its low computational burden and compact object-based input representation, the approach has also been deployed as an educational nowcasting challenge at IndabaX Deep Learning Madagascar 2024 (IndabaX Madagascar, 2024), illustrating its accessibility for capacity building in resource-constrained environments.

The domain-wide NCAST model remains computationally tractable at regional to continental scales. With approximately nine million parameters, training can be performed efficiently on a single modern (A100) GPU, while inference, including deep-ensemble configurations, can be executed on standard CPU hardware. Importantly, inference cost scales primarily with the number of detected convective cores rather than with domain size or image resolution. This makes NCAST substantially more efficient than fully image-based deep-learning approaches when applied over large regions such as the western Sahel or pan-African domains.

The efficiency advantage is particularly evident when compared with image-based transformer architectures such as Vision Transformers (ViTs). In the present configuration, NCAST processes at most approximately 250 object tokens per sample, whereas a ViT operating on five (512 × 512) images with a (16 × 16) patch size would process approximately 5120 tokens in a representative configuration. This leads to substantially higher computational and memory costs for image-based approaches. Moreover, pixel-based architectures must learn convective features directly from raw fields, typically requiring larger training datasets and longer optimisation schedules.

The framework has been extended to a pan-African configuration (Pan-African NCAST; PANCAST) and is currently deployed through the UKCEH nowcasting portal (UK Centre for Ecology & Hydrology, 2024). The complete pipeline, including core detection, feature extraction, preprocessing, and inference for all lead times up to 2 h, is typically completed within approximately five minutes of receiving the satellite observations. PANCAST has also been evaluated within operational forecasting environments, including the EW4Energy testbed (Department for Science, Innovation and Technology, 2025; FASTA, 2026).

The principal limitation of the object-based formulation is its reliance on the convective-core detection algorithm used to define the input representation. In the present study, cores are identified using cold cloud-top temperature thresholds and wavelet filtering; consequently, warm-rain and shallow convective systems are not explicitly targeted. The model is applied only in situations where at least one convective core is detected within the input time window and is therefore designed to forecast the subsequent evolution and organisation of existing convection rather than isolated first initiation events. Convective initiation triggered by interactions between existing convective cores may still be captured indirectly. However, mesoscale boundaries and environmental controls such as wind shear, moisture, and surface conditions are not included explicitly as predictors and can influence the forecast only through their effects on the detected convective-core population up to the nowcast origin time.

Only marginal differences were observed between the temporal-history configurations. This may indicate that much of the predictive information is already contained in the current convective-core configuration. However, it may also reflect architectural choices within NCAST. Convective cores from different observation times are processed jointly as a common set of tokens, and temporal information is aggregated through pooling rather than being represented as an explicit ordered sequence. Consequently, additional historical cores may not be fully exploited within the current architecture.

In addition, mapping the pooled latent object-based representation produced by the self-attention mechanism onto a spatial grid introduces an explicit object-to-field projection step that purely image-based models do not require. While this projection is necessary to generate gridded probabilistic outputs, alternative formulations operating directly on object-to-object interactions, such as graph-based representations, may provide a more natural treatment of organised convective systems and their evolution. A promising direction for future work would be to represent tracked convective structures as explicit objects, allowing their growth, propagation, merging, and decay to be modelled directly through time rather than inferred from a pooled set of convective-core observations.

Despite these limitations, the results demonstrate that core-level information alone contains substantial predictive skill for short-range convective nowcasting. The object-based design remains data-efficient, computationally scalable, and operationally deployable, making it particularly well suited to real-time applications over large domains and in operational environments where computational efficiency is important.

6 Conclusions

This study has developed and evaluated an object-based deep learning framework for probabilistic nowcasting of convective cores using compact descriptors derived from geostationary satellite observations.

At a single location, core-level attributes—particularly proximity, size, and intensity—are shown to contain sufficient information to support skilful probabilistic forecasts up to 6 hours ahead, outperforming an operational conditional-climatology benchmark. Explainability analyses and controlled spatial-transfer experiments indicate that the learned relationships are physically coherent: short-range skill is governed primarily by proximity and persistence, whereas longer lead times increasingly reflect modulation by the broader diurnal and seasonal environment associated with the West African monsoon.

When extended to the western Sahel, the NCAST framework represents convective-core populations using a spatio-temporal Transformer architecture and produces coherent gridded probabilistic forecasts. Across independent multi-year evaluation, the model maintains strong discrimination and positive probabilistic skill relative to both persistence and conditional climatology, while also demonstrating meaningful spatial skill at neighbourhood scales relevant to organised convection. Ablation experiments further demonstrate that self-attention provides the largest contribution to forecast skill, highlighting the importance of representing interactions among convective cores. Temporal features also become increasingly important at longer lead times.

More broadly, the findings suggest that skilful convective-core nowcasts can be derived directly from the evolving population of convective cores, even in data-sparse environments. This provides a foundation for future nowcasting systems that integrate object-based representations with additional environmental and Earth-observation information to further improve forecast skill.

7 Author contributions

Mendrika Rakotomanga: Conceptualisation; Methodology; Software; Validation; Formal analysis; Investigation; Data curation; Visualisation; Writing—original draft; Writing—review & editing.

Douglas J. Parker: Conceptualisation; Methodology; Supervision; Funding acquisition; Writing—review & editing.

Nadhir Ben Rached: Conceptualisation; Methodology; Supervision; Writing—review & editing.

Cornelia Klein: Conceptualisation; Methodology; Supervision; Writing—review & editing.

Seonaid R. Anderson: Conceptualisation; Methodology; Supervision; Writing—review & editing.

Steven C. Wells: Software; Data curation; Writing—review & editing.

8 Funding

This work was supported by the Faculty of Engineering and Physical Sciences Sub-Saharan Africa Scholarship at the University of Leeds. Additional support was provided through the WISER Early Warnings for Southern Africa (WISER-EWSA) project, funded by the Met Office as part of the Weather and Climate Information Services (WISER) programme on behalf of the UK Government's Foreign, Commonwealth and Development Office (FCDO); the UK Centre for Ecology & Hydrology NC–International Visiting Scientist scheme; and the Higher Education Innovation Fund (HEIF), University of Leeds.

9 Acknowledgements

The authors thank the NFLICS project team for providing the conditional-climatology benchmark used in this study, and Jawairia Ahmad, Christopher Taylor, Kilian Hermes and Charles Taylor for valuable discussions and technical support.

10 Data availability

Cloud-top temperature (CTT) data from Meteosat Second Generation (MSG) SEVIRI (IR 10.8 µm channel) are available from EUMETSAT (`https://www.eumetsat.int/`) subject to standard access conditions.

The convective-core identification algorithm is publicly available at `https://github.com/cornkle/ccores`. The convective-core database and NFLICS conditional-climatology fields are maintained by the UK Centre for Ecology & Hydrology and are available upon reasonable request, subject to data-sharing agreements.

11 Code availability and software

Computations were performed on JASMIN, the UK national high-performance computing facility for environmental data analysis. The models and evaluation pipelines were implemented in Python using PyTorch and related scientific libraries. The code required to reproduce the experiments is available from the authors upon reasonable request and will be archived in a public repository upon publication.

12 Conflict of Interest

The authors declare no conflict of interest.

13 References

Adler, R. F. & Mack, R. A. (1986). “Thunderstorm cloud top dynamics as inferred from satellite observations and a cloud top parcel model”. In: *Journal of the Atmospheric Sciences* 43.18, pp. 1945–1960. DOI: 10.1175/1520-0469(1986)043<1945:TCTDAI>2.0.CO;2.

Agence Nationale de la Statistique et de la Démographie (2023). *Situation Économique et Sociale Régionale – Dakar*. Accessed 4 August 2025. URL: https://www.ansd.sn/themes/situation-economiques-et-sociales/ses-regionales?field_year_value=All&field_regions_ses_target_id=5.

Anderson, S. R., Cole, S. J., Klein, C., Taylor, C. M., Diop, C. A. & Kamara, M. (2024). “Nowcasting convective activity for the Sahel: A simple probabilistic approach using real-time and historical satellite data on cloud-top temperature”. In: *Quarterly Journal of the Royal Meteorological Society* 150.759, pp. 597–617. DOI: 10.1002/qj.4607.

Bassford, J., Parker, D. J., Marsham, J. H. & Warner, J. (2025). “Scale Interactions of Deep Convection in the Sahel: Exploring Ring Patterns in Lagged Cross-Correlations of Rainfall”. In: *Quarterly Journal of the Royal Meteorological Society*, e70041. DOI: 10.1002/qj.70041.

Billet, J., DeLisi, M., Smith, B. G. & Gates, C. (1997). “Use of regression techniques to predict hail size and the probability of large hail”. In: *Weather and Forecasting* 12.1, pp. 154–164. DOI: 10.1175/1520-0434(1997)012<0154:UORTTP>2.0.CO;2.

Birch, C. E., Parker, D. J., O’Leary, A., Marsham, J. H., Taylor, C. M., Harris, P. P. & Lister, G. M. S. (2013). “Impact of Soil Moisture and Convectively Generated Waves on the Initiation of a West African Mesoscale Convective System”. In: *Quarterly Journal of the Royal Meteorological Society* 139.676, pp. 1712–1729. DOI: 10.1002/qj.2062.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Breiman, L. (2001). “Random Forests”. In: *Machine Learning* 45.1, pp. 5–32. DOI: 10.1023/A:1010933404324. URL: https://doi.org/10.1023/A:1010933404324.

Brier, G. W. (1950). “Verification of Forecasts Expressed in Terms of Probability”. In: *Monthly Weather Review* 78.1, pp. 1–3. DOI: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

Burton, R. R., Blyth, A. M., Cui, Z., Groves, J. L., Lamptey, B. L., Fletcher, J. K., Marsham, J. H., Parker, D. J. & Roberts, A. J. (2022). “Satellite-Based Nowcasting of West African Mesoscale Storms Has Skill at up to Four Hours Lead Time”. In: *Weather and Forecasting* 37.4, pp. 445–455. DOI: 10.1175/WAF-D-21-0051.1.

Chen, T. & Guestrin, C. (2016). “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. DOI: 10.1145/2939672.2939785. URL: https://doi.org/10.1145/2939672.2939785.

Cintineo, J. L., Pavolonis, M. J., Sieglaff, J. M., Wimmers, A. J., Brunner, J. C. & Bellon, W. (2020). “A deep-learning model for automated detection of intense midlatitude convection using geostationary satellite images”. In: *Weather and Forecasting* 35.6, pp. 2567–2588. DOI: 10.1175/WAF-D-20-0028.1.

Cortes, C. & Vapnik, V. (1995). “Support-Vector Networks”. In: *Machine Learning* 20.3, pp. 273–297. DOI: 10.1007/BF00994018. URL: https://doi.org/10.1007/BF00994018.

Cox, D. R. (1958). “The Regression Analysis of Binary Sequences”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 20.2, pp. 215–242. URL: http://www.jstor.org/stable/2983890.

Dai, K., Li, X., Fang, J., Ye, Y., Yu, D., Su, H., Xian, D., Qin, D. & Wang, J. (2025). “Four-hour thunderstorm nowcasting using a deep diffusion model for satellite data”. In: *Proceedings of the National Academy of Sciences* 122.51, e2517520122. DOI: 10.1073/pnas.2517520122.

Department for Science, Innovation and Technology (2025). *Developing Capacity for Storm and Lightning Early Warning for the Energy Sector in Ghana (EW4Energy)*. Accessed 16 June 2026. URL: https://devtracker.fcdo.gov.uk/programme/GB-GOV-26-ISPF-STFC-DQ5ZR34-KMC3QB9-B4PNUPA/summary.

Dixon, M. & Wiener, G. (1993). “TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A Radar-Based Methodology”. In: *Journal of Atmospheric and Oceanic Technology* 10.6, pp. 785–797. DOI: 10.1175/1520-0426(1993)010<0785:TTITAA>2.0.CO;2.

EUMETSAT (2022). *EUMETSAT User Portal*. https://www.eumetsat.int/eumetsat-user-portal. Accessed: 2026-02-25.

FASTA (2026). *EW4Energy Testbed Portal*. Accessed 16 June 2026. URL: https://fastaweather.com/testbed-portal/.

Fawcett, T. (2006). “An introduction to ROC analysis”. In: *Pattern Recognition Letters* 27.8, pp. 861–874. DOI: 10.1016/j.patrec.2005.10.010.

Gagne, D. J., McGovern, A. & Brotzge, J. (2009). “Classification of convective areas using decision trees”. In: *Journal of Atmospheric and Oceanic Technology* 26.7, pp. 1341–1353. DOI: 10.1175/2008JTECHA1205.1.

Gijben, M. & De Coning, E. (2017). “Using Satellite and Lightning Data to Track Rapidly Developing Thunderstorms in Data-Sparse Regions of Africa”. In: *Atmosphere* 8.4, p. 67. DOI: 10.3390/atmos8040067.

Handwerker, J. (2002). “Cell tracking with TRACE3D—a new algorithm”. In: *Atmospheric Research* 61.1, pp. 15–34. DOI: 10.1016/S0169-8095(01)00100-4.

Hermes, K., Marsham, J. H., Bollasina, M., Brooks, M., Klose, M. & Marenco, F. (2025). "Nowcasting of Dust and Convective Storms via Diffusion-Model Predictions of SEVIRI RGB Imagery". In: *Weather and Climate Extremes* 50, p. 100828. DOI: 10.1016/j.wace.2025.100828.

Hill, P. G., Stein, T. H. M., Roberts, A. J., Fletcher, J. K., Marsham, J. H. & Groves, J. (2020). "How skilful are Nowcasting Satellite Applications Facility products for tropical Africa?" In: *Meteorological Applications* 27.6, e1966. DOI: 10.1002/met.1966.

Hochreiter, S. & Schmidhuber, J. (1997). "Long short-term memory". In: *Neural Computation* 9.8, pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

IndabaX Madagascar (2024). *Storm Nowcasting Hackathon – IndabaX Madagascar 24*. Kaggle competition page. URL: https://kaggle.com/competitions/hackathon-DL-IndabaX-MDG-2024.

Ioffe, S. & Szegedy, C. (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *Proceedings of the 32nd International Conference on Machine Learning*, pp. 448–456.

Keane, R. J., Parker, D. J., Dunn-Sigouin, E., Kolstad, E. W. & Marsham, J. H. (2025). "Mid-Latitude Versus Tropical Scales of Predictability and Their Implications for Forecasting". In: *Meteorological Applications* 32.4, e70055. DOI: 10.1002/met.70055.

Kingma, D. P. & Ba, J. (2015). "Adam: A method for stochastic optimization". In: *International Conference on Learning Representations*.

Kitzmiller, D. H., McGovern, W. E. & Saffle, R. F. (1995). "The WSR-88D severe weather potential algorithm". In: *Weather and Forecasting* 10.1, pp. 141–159. DOI: 10.1175/1520-0434(1995)010<0141:TWSWPA>2.0.CO;2.

Klein, C., Belušić, D. & Taylor, C. M. (2018). "Wavelet scale analysis of mesoscale convective systems for detecting deep convection from infrared imagery". In: *Journal of Geophysical Research: Atmospheres* 123.6, pp. 3035–3050. DOI: 10.1002/2017JD027432.

Klein, C., Nkrumah, F., Taylor, C. M. & Adefisan, E. A. (2021). "Seasonality and trends of drivers of mesoscale convective systems in southern West Africa". In: *Journal of Climate* 34.1, pp. 71–87. DOI: 10.1175/JCLI-D-20-0194.1.

Kniffka, A., Knippertz, P., Fink, A. H., Benedetti, A., Brooks, M., Hill, P. G., Maranan, M., Pante, G. & Vogel, B. (2020). "An evaluation of operational and research weather forecasts for southern West Africa using observations from the DACCIWA field campaign in June–July 2016". In: *Quarterly Journal of the Royal Meteorological Society* 146.727, pp. 2131–2155. DOI: 10.1002/qj.3729.

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". In: *Monthly Weather Review* 149.12, pp. 3897–3921. DOI: 10.1175/MWR-D-21-0096.1.

Laing, A. G. & Fritsch, J. M. (1993). "Mesoscale convective complexes in Africa". In: *Monthly Weather Review* 121.8, pp. 2254–2267. DOI: 10.1175/1520-0493(1993)121<2254:MCCIA>2.0.CO;2.

Laing, A. G. & Fritsch, J. M. (2000). "The large-scale environments of the global populations of mesoscale convective complexes". In: *Monthly Weather Review* 128.8, pp. 2756–2776. DOI: 10.1175/1520-0493(2000)128<2756:TLSEOT>2.0.CO;2.

LeCun, Y., Bottou, L., Orr, G. B. & Müller, K.-R. (2012). "Efficient BackProp". In: *Neural Networks: Tricks of the Trade*. Ed. by Montavon, G., Orr, G. B. & Müller, K.-R. Vol. 7700. Lecture Notes in Computer Science. Springer, pp. 9–48. DOI: 10.1007/978-3-642-35289-8_3.

Leinonen, J., Hamann, U., Sideris, I. V. & Germann, U. (2023). "Thunderstorm nowcasting with deep learning: A multi-hazard data fusion model". In: *Geophysical Research Letters* 50.8, e2022GL101626. DOI: 10.1029/2022GL101626.

Loshchilov, I. & Hutter, F. (2019). "Decoupled Weight Decay Regularization". In: *arXiv preprint arXiv:1711.05101*. DOI: 10.48550/arXiv.1711.05101.

Lundberg, S. (2018). *SHAP (SHapley Additive exPlanations)*. Accessed 8 August 2025. URL: https://github.com/shap/shap.

Lundberg, S. M. & Lee, S.-I. (2017). "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems 30*, pp. 4765–4774.

Maidment, R. I., Grimes, D., Allan, R. P., Tarnavsky, E., Stringer, M., Hewison, T., Roebeling, R. & Black, E. (2014). "The 30 year TAMSAT African rainfall climatology and time series (TARCAT) data set". In: *Journal of Geophysical Research: Atmospheres* 119.18, pp. 10619–10644. DOI: 10.1002/2014JD021927.

Maranan, M., Fink, A. H. & Knippertz, P. (2018). "Rainfall types over southern West Africa: Objective identification, climatology and synoptic environment". In: *Quarterly Journal of the Royal Meteorological Society* 144.714, pp. 1628–1648. DOI: 10.1002/qj.3345.

Markowski, P. & Richardson, Y. (2010). *Mesoscale Meteorology in Midlatitudes*. Chichester: John Wiley & Sons. ISBN: 9780470742136.

Marr, D. & Hildreth, E. (1980). "Theory of edge detection". In: *Proceedings of the Royal Society of London. Series B. Biological Sciences* 207.1167, pp. 187–217. DOI: 10.1098/rspb.1980.0020.

Mathon, V., Laurent, H. & Lebel, T. (2002). “Mesoscale convective system rainfall in the Sahel”. In: *Journal of Applied Meteorology* 41.11, pp. 1081–1092. DOI: 10.1175/1520-0450(2002)041<1081:MCSRIT>2.0.CO;2.

Mecikalski, J. R., Sandmæl, T. N., Murillo, E. M., Homeyer, C. R., Bedka, K. M., Apke, J. M. & Jewett, C. P. (2021). “A random-forest model to assess predictor importance and nowcast severe storms using high-resolution radar–GOES satellite–lightning observations”. In: *Monthly Weather Review* 149.6, pp. 1725–1746. DOI: 10.1175/MWR-D-19-0274.1.

Nair, V. & Hinton, G. E. (2010). “Rectified linear units improve restricted Boltzmann machines”. In: *Proceedings of the 27th International Conference on Machine Learning*, pp. 807–814.

Nicholson, S. E. (2009). “A Revised Picture of the Structure of the Monsoon and Land ITCZ over West Africa”. In: *Climate Dynamics* 32, pp. 1155–1171. DOI: 10.1007/s00382-008-0514-3.

Panthou, G., Vischel, T., Lebel, T., Quantin, G., Sane, Y., Ba, A., Ndiaye, O., Diongue-Niang, A. & Diopkane, M. (2018). “Rainfall intensification in tropical semi-arid regions: The Sahelian case”. In: *Environmental Research Letters* 13.6, p. 064013. DOI: 10.1088/1748-9326/aac334.

Parker, D. J., Thorncroft, C. D., Burton, R. R. & Diongue-Niang, A. (2005). “Analysis of the African easterly jet, using aircraft observations from the JET2000 experiment”. In: *Quarterly Journal of the Royal Meteorological Society* 131.608, pp. 1461–1482. DOI: 10.1256/qj.03.189.

Pulkkinen, S., Nerini, D., Pérez Hortal, A. A., Velasco-Forero, C., Seed, A., Germann, U. & Foresti, L. (2019). “Pysteps: An open-source Python library for probabilistic precipitation nowcasting (v1.0)”. In: *Geoscientific Model Development* 12.10, pp. 4185–4219. DOI: 10.5194/gmd-12-4185-2019.

Roberts, A. J., Fletcher, J. K., Groves, J., Marsham, J. H., Parker, D. J., Blyth, A. M., Adefisan, E. A., Ajayi, V. O., Barrette, R., Coning, E. de, Dione, C., Diop, A., Foamouhoue, A. K., Gijben, M., Hill, P. G., Lawal, K. A., Mutemi, J., Padi, M., Popoola, T. I., Rípodas, P., Stein, T. H. M. & Woodhams, B. J. (2022). “Nowcasting for Africa: advances, potential and value”. In: *Weather* 77.7, pp. 250–256. DOI: 10.1002/wea.3936.

Roberts, N. M. & Lean, H. W. (2008). “Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events”. In: *Monthly Weather Review* 136.1, pp. 78–97. DOI: 10.1175/2007MWR2123.1.

Shapley, L. S. (1953). “A Value for n-Person Games”. In: *Contributions to the Theory of Games II*. Ed. by Kuhn, H. W. & Tucker, A. W. Annals of Mathematics Studies 28. Princeton University Press, pp. 307–317.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014). “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *Journal of Machine Learning Research* 15.56, pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

Taylor, C. M., Belušić, D., Guichard, F., Parker, D. J., Vischel, T., Bock, O., Harris, P. P., Janicot, S., Klein, C. & Panthou, G. (2017). “Frequency of extreme Sahelian storms tripled since 1982 in satellite observations”. In: *Nature* 544.7651, pp. 475–478. DOI: 10.1038/nature22069.

Taylor, C. M., Klein, C., Dione, C., Parker, D. J., Marsham, J. H., Diop, C. A., Fletcher, J. K., Chaibou, A. A. S., Nafissa, D. B., Semeena, V. S., Cole, S. J. & Anderson, S. R. (2022). “Nowcasting tracks of severe convective storms in West Africa from observations of land surface state”. In: *Environmental Research Letters* 17.3, p. 034016. DOI: 10.1088/1748-9326/ac536d.

UK Centre for Ecology & Hydrology (2024). *About the Nowcasting Portal*. https://africa-hydrology.ceh.ac.uk/nowcasting/about. Accessed 24 July 2024.

UKRI Gateway to Research (2025). *Nowcasting Flood Impacts of Convective Storms in the Sahel (NFLICS)*. Accessed 23 February 2026. URL: https://gtr.ukri.org/projects?ref=NE%2FS006087%2F1.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). “Attention Is All You Need”. In: *Advances in Neural Information Processing Systems 30*, pp. 6000–6010.

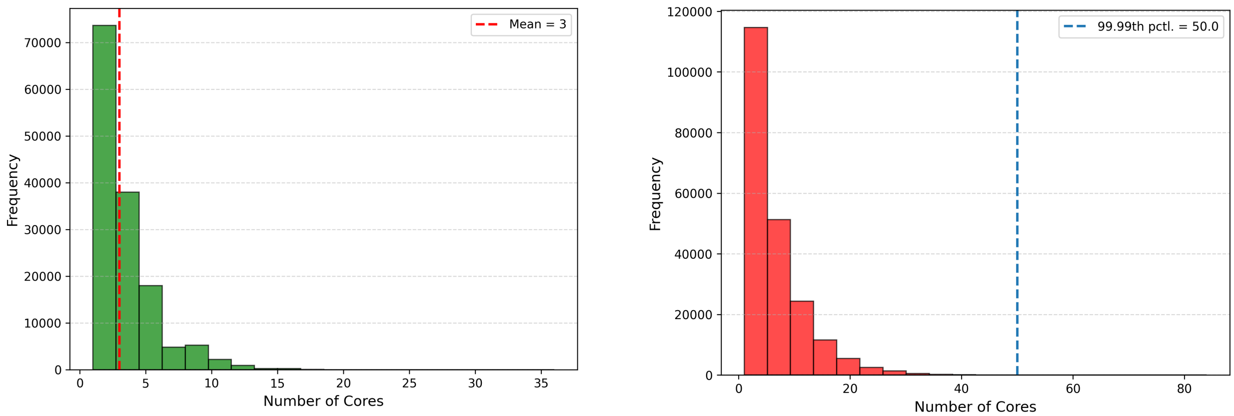
Vogel, P., Knippertz, P., Fink, A. H., Schlueter, A. & Gneiting, T. (2018). “Skill of global raw and postprocessed ensemble predictions of rainfall over northern tropical Africa”. In: *Weather and Forecasting* 33.2, pp. 369–388. DOI: 10.1175/WAF-D-17-0127.1.

World Meteorological Organization (2017). *Guidelines for Nowcasting Techniques*. WMO-No. 1198. Available at: https://library.wmo.int/idurl/4/55666.

– (2025). *WMO Radar Database*. Accessed 16 December 2025. URL: https://wrd.mgm.gov.tr/Home/Wrd.

Zipser, E. J., Cecil, D. J., Liu, C., Nesbitt, S. W. & Yorty, D. P. (2006). “Where are the most intense thunderstorms on Earth?” In: *Bulletin of the American Meteorological Society* 87.8, pp. 1057–1071. DOI: 10.1175/BAMS-87-8-1057.

A Core count distribution



(a) (b)

Figure A1: Distribution of detected convective core counts at 15-minute intervals during 2004–2024, conditioned on the presence of at least one core. Panel (a) shows the $12^{\circ} \times 12^{\circ}$ Dakar context domain, and panel (b) the western Sahel (WS) domain. The dashed lines indicate the mean (a) and the 99.99th percentile (b), respectively.

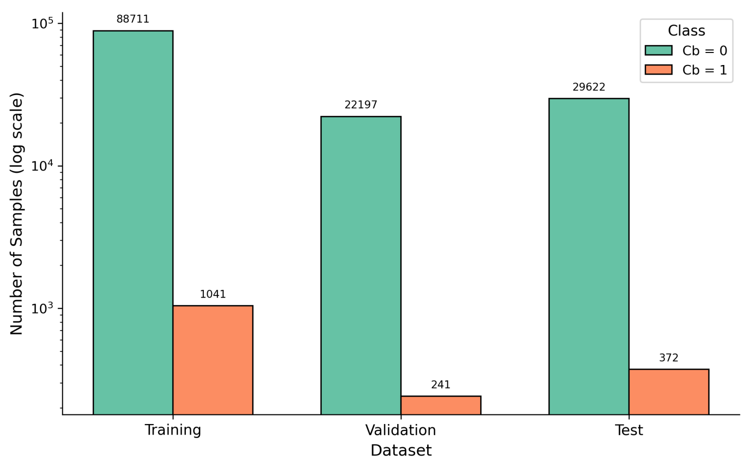


Figure A2: Class distribution across the training, validation, and test sets for the 1-hour lead time in Dakar. Bars show the number of samples with ($y_{t_0+1\mathrm{h}} = 1$) and without ($y_{t_0+1\mathrm{h}} = 0$) convective-core occurrence. The pronounced class imbalance highlights the difficulty of standard binary classification metrics.

B Hyperparameter Configurations for Single-Location Models

Model	Lead time	Best configuration
LR	1 h	$C = 0.01$
	3 h	$C = 0.01$, class weight=balanced
	6 h	$C = 0.01$, class weight=balanced
SVM	1 h	$C = 100$, class weight=balanced
	3 h	$C = 100$, class weight=balanced
	6 h	$C = 0.1$
XGBoost	1 h	$n = 500$, depth=5, lr=0.01, subsample=1.0, colsample=1.0
	3 h	$n = 100$, depth=5, lr=0.05, subsample=0.8, colsample=0.8
	6 h	$n = 100$, depth=7, lr=0.01, subsample=1.0, colsample=0.8
RF	1 h	$n = 100$, depth=None, min leaf=2
	3 h	$n = 100$, depth=8, min leaf=2
	6 h	$n = 500$, depth=12, min leaf=1, class weight=balanced

Table 4: Best hyperparameter configurations for the single-location benchmark machine-learning models at forecast lead times of 1, 3, and 6 h, selected using validation AUC. LR denotes logistic regression, SVM support vector machine, RF random forest, and XGBoost Extreme Gradient Boosting. For LR and SVM, C denotes the inverse regularisation strength. For XGBoost, n denotes the number of trees, depth the maximum tree depth, lr the learning rate, subsample the fraction of training samples used for each tree, and colsample the fraction of features used for each tree. For RF, n denotes the number of trees, depth the maximum tree depth, and min leaf the minimum number of samples required in a leaf node. The setting class weight=balanced applies class weights inversely proportional to class frequencies in the training data.