Weather and Climate Extremes 50 (2025) 100828

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Nowcasting of dust and convective storms via diffusion-model predictions of SEVIRI RGB imagery

Kilian Hermes [a,*], John H. Marsham [a], Massimo Bollasina [b], Melissa Brooks [c], Martina Klose [d], Franco Marenco [e]

[a] University of Leeds, Institute for Climate and Atmospheric Science, Leeds, United Kingdom
[b] School of GeoSciences, University of Edinburgh, Edinburgh, United Kingdom
[c] Met Office, Exeter, United Kingdom
[d] Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research, Karlsruhe, Germany
[e] The Cyprus Institute, Nicosia, Cyprus

ARTICLE INFO

Keywords:
Nowcasting
Dust storms
Diffusion model
Machine learning

ABSTRACT

Convective storms in the tropics are inherently unpredictable on the scales typical of global Numerical Weather Prediction (NWP) models. Rapid-refresh short term predictions, so called nowcasts, provide added value and can improve forecasts on short timescales. Nowcasts provide the most up to date predictions, making them particularly relevant for rapidly developing high impact weather that is not reproduced in global NWP models. While NWP nowcasts exist for USA, UK, and continental Europe, this is not the case for Africa where nowcasts are primarily observation-based. Here we focus on West Africa, an area where convective storms are frequent weather events. Besides direct impacts from the convective storms, outflow from these storms frequently causes large-scale dust storms. Dust storms are high impact weather and directly impact human life. Early warning is crucial for mitigating their adverse impacts. While dust storms are poorly forecast in currently operational weather prediction models, they are well observed from space, posing great potential for nowcasting. The desert dust red–green–blue (RGB) composite highlights dust and convective systems in bright colours, making it a useful product for a trained observer for identifying and tracking dust storms. In this study, we introduce DustCast, a diffusion model for image-based nowcasting of dust storms by predicting the SEVIRI desert dust RGB composite up to 6 h ahead. Our nowcasts can reproduce convectively generated dust storms that currently operational NWP do not reliably reproduce. We also predict convective storms that are contained in the same imagery and provide useful context information for a forecaster assessing the weather situation. Our model shows limited capability of reproducing entirely new features that are not contained in the input data. This primarily poses limitations for cases with convection initiation. On average, our model achieves useful skill (Fractions Skill Score > 0.5) for predicting dust storms up to 5 h lead time, and for convective systems for up to 4 h. DustCast is the first model of its kind for nowcasting dust, and extends skill for nowcasting convective storms by more than 2 h compared to conventional methods based on optical flow. Deployment during a nowcasting testbed has shown that our nowcast provides an easy-to-use product for operational forecasters. Our method could also be adapted to predict other RGB composites such as those specifically for convection, ash or fog, and indeed other products using observation data from geostationary satellites, opening potential for a large variety of applications.

1. Introduction

Numerical weather prediction (NWP) models are a well established tool for predicting the weather around the globe. While NWP models perform well for most situations, observation-based short-term predictions (nowcasts) provide a complementary approach on short timescales. Nowcasts can be frequently initialised and provide the most up-to-date forecasts for short timescales with typical lead times of up to 6 h (Burton et al., 2022; Roberts et al., 2022). Although the World Meteorological Organisation (WMO) definition of nowcasts includes short-range predictions from NWP, generally a large focus is on observation-driven predictions (World Meteorological Organization, 2017). While rapid-refresh convection-permitting NWP is available for

* Corresponding author.
E-mail address: kilian.hermes@e.mail.de (K. Hermes).

https://doi.org/10.1016/j.wace.2025.100828
Received 29 August 2025; Received in revised form 3 November 2025; Accepted 3 November 2025
Available online 14 November 2025

USA, UK and mainland Europe, such forecasts are currently unavailable for Africa. Therefore, nowcasting in Africa is essentially based on observations (Roberts et al., 2022). Such nowcasts are particularly valuable when they capture features that NWP has missed or mis-represented, and allow warnings for fast-developing high-impact weather. Storms in the tropics are inherently unpredictable on NWP timescales (Keane et al., 2025), underlining the added value from nowcasts.

Here we focus on West Africa, a region that is frequently affected by convective storms and convectively-generated dust storms (Knippertz et al., 2007; Knippertz and Todd, 2010). Dust storms are high-impact events (Knippertz et al., 2007; Knippertz and Todd, 2010) and directly impact human life (Middleton, 2017) by disrupting land (Bhattachan et al., 2019) and air traffic (Baddock et al., 2013), affect energy delivery from solar energy systems (Sarver et al., 2013), pose a health threat (de Longueville et al., 2013) and cause significant financial loss (Al-Hemoud et al., 2017). Early predictions of dust storms are crucial for mitigating adverse impacts. In West Africa, cold pools from moist convective systems frequently cause particularly large-scale dust storms (Trzeciak et al., 2017). Also known as haboobs, these storms can travel more than 1,000 km (Flamant et al., 2007; Roberts and Knippertz, 2014), and their occurrence is closely linked to moist convective activity (Heinold et al., 2013). While currently operational NWP models and reanalyses perform well in reproducing mean flow and surface winds during winter when convective activity is low, cold pools which are frequent during summer are often missed due to moist convection being parametrised (Marsham et al., 2011; Roberts et al., 2017; Trzeciak et al., 2017). Within convection parameterisations convective up and down draughts are assumed to occur within a column, and resultant cold pools are very weak or missing. Further uncertainties are introduced through the non-linear nature of dust emission and their parameterisations, resulting in many large dust storms being missed in currently operational models (Marsham et al., 2011; Heinold et al., 2013; Trzeciak et al., 2017). While radar coverage is limited and in-situ observations in West Africa are sparse (Cowie et al., 2014), instruments based on geostationary satellites provide excellent data coverage and potential for nowcasting (Browning and Collier, 1989; Roberts et al., 2022).

Traditional optical-flow-based (OF) nowcasts extract motion from a sequence of observations and generate predictions by forward extrapolating pixel locations based on these motion vectors. Such forward extrapolations typically start from recent observations, and therefore only predict features that are already contained in the observations. While OF nowcasts for precipitation and storms have successfully been deployed for more than two decades (e.g., Turner et al., 2004; Bowler et al., 2006; Srivastava et al., 2012), nowcasting the initiation of convection remains challenging (Smith et al., 2024). Traditional OF nowcasts further reach limits for scenes with occluded features, or scenes with colour and brightness changes in the input frames. In such situations, detecting the relevant flow is challenging (Alfarano et al., 2024), while forward extrapolations can suffer from additional issues such as scenes with opposing motion vectors that can unrealistically distort the data field. For observations of dust, it is common that high or mid-level clouds move in different directions than the dust which is close to the surface, and that dust is partially hidden under clouds. This is where machine learning (ML) could become particularly valuable. ML excels for modelling non-linear systems, as it can capture complex relationships directly from data. ML can reliably derive multi-scale motion directly from sequences of images without the explicit assumptions that are required for OF methods (Ilg et al., 2017). ML models have been shown capable of predicting images with moving, colliding, merging or occluded features (e.g., Denton and Fergus, 2018; Wang et al., 2018), a task that resembles weather observations from a satellite.

Recent advances in ML methods have led to their adaptation for weather prediction purposes (Keisler, 2022; Bi et al., 2022; Lam et al., 2023; Lang et al., 2024). However, as of 2025, most ML-based weather prediction models (MLWP) are trained on reanalyses and can therefore predict only features that are contained in the model analyses. Therefore, also reanalysis-trained MLWP cannot reliably predict dust storms.

While these issues make dust storm forecasts in NWP and MLWP challenging, these storms are well observed from space. The Spinning Enhanced Visible and Infrared Instrument (SEVIRI) onboard the Meteosat Second Generation (MSG) satellites provides measurements on 12 spectral channels from the visible range to the infrared. The red–green–blue (RGB) false-colour-composite for desert dust (Lensky and Rosenfeld, 2008) highlights dust in bright colours. A trained observer can identify and track storms from the dust composite images and their evolution in time which makes it a very useful tool for studying dust storms in North Africa. This composite has been used for various applications, such as to identify dust source regions and their frequency of activation (Schepanski et al., 2007; Ashpole and Washington, 2013), assess dust transport (Knippertz and Todd, 2012), and is used for operational dust forecasting. ML nowcasts for the RGB composites that highlight convective clouds (Ionescu et al., 2021) and fog (Bari et al., 2023) have been developed on country-scale domains. These methods can provide skilful forecasts on short lead times, but are limited by their domain size through missing forcings that may be advected into the domain through its boundaries and suffer from blurring with increasing lead times. Blurring is a common issue of ML predictions when optimised with loss functions such as mean squared error (MSE). This can be seen as a visualisation of forecast uncertainty, however limits the interpretability for the human forecaster and does not allow comprehensive probabilistic predictions through scenario assessment.

Here we use a diffusion model (DM) (Sohl-Dickstein et al., 2015) to generate nowcasts. DMs are a family of deep generative models and achieve state-of-the-art performance in image generation tasks (for a comprehensive overview see Yang et al., 2023). DMs use stochastic differential equations to model an iterative forward process that incrementally destroys data by injecting noise and a backward process in which the models learn to stepwise reverse this process for generating new samples from noise. This process can be conditioned, e.g., with additional image input. Furthermore, DMs have an ensemble method built into their architecture. Sampling from different noise fields results in different but sharp new realisations. DMs have been shown to outperform previous methods in terms of image quality and ensemble diversity even if trained using a MSE loss (Dhariwal and Nichol, 2021) and have been adapted for nowcasting precipitation (Leinonen et al., 2023), hail (Shi et al., 2025), or convective clouds (Chen et al., 2025).

In this study, we introduce DustCast, a DM for image-based nowcasting of false-colour RGB dust imagery for a large domain over West Africa up to 6 h ahead. By nowcasting images, DustCast also nowcasts clouds from convective storms that are contained in the same imagery. Section 2 provides an overview about data, generation of the dust composite imagery and data-preprocessing for machine learning, Section 3 summarises the implementation of the nowcasting model and explains verification metrics, Section 4 evaluates the model performance on unseen data, and Section 5 summarises conclusions and provides suggestions for future model improvements and applications.

2. Data and methods

2.1. SEVIRI observation data

We use brightness temperature (BT) from SEVIRI onboard MSG geostationary satellites positioned at 0 °N, 0 °E, operated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). SEVIRI provides retrievals on twelve spectral channels from the visible to infrared (0.4–13.4 μm), with a spatial resolution of 1 to 3 km at nadir, and a temporal resolution of 15 min for a full-disk scan (Schmetz et al., 2002). The data record covers more than 20 years and new data are available in near real time directly

Table 1
Colour channel and assigned Brightness Temperate (BT) or Brightness Temperature Differences (BTD) from SEVIRI, scaling range and gamma correction factor for the desert dust RGB composite[a].

Colour	Channel	BT_{min}(K)	BT_{max}(K)	Γ
Red	BTD(120–108)	−4	2	1
Green	BTD(108–087)	0	15	2.5
Blue	BT(108)	261	289	1

[a] After Lensky and Rosenfeld (2008).

Table 2
Criteria[a] for detecting dust in the SEVIRI RGB composite from thresholds for Brightness Temperate (BT) and Brightness Temperature Differences (BTD).

Colour	Channel	Criterion
Red	BTD(120–108)	≥ 285 K
Green	BTD(108–087)	≥ 0 K
Blue	BT(108)	≤ 10 K
Green anomaly	$BTD(108–087)_{anom}$	≤ −2 K

[a] After Ashpole and Washington (2012).

from EUMETSAT. This high temporal resolution and low latency makes SEVIRI BT data ideal for nowcasting purposes. For this study, we use Level 1.5 data (EUMETSAT, 2017) from the channels centred at 8.7, 10.8, and 12.0 μm (the same channels that are exploited in the desert dust RGB composite). These data come geolocated and radiometrically calibrated and are provided in a geostationary projection with 3 km resolution at the sub-satellite point. We further use the Meteosat cloud mask (EO:EUM:DAT:MSG:CLM, EUMETSAT, 2015) for cloud screening.

2.2. Desert dust RGB composite

The RGB composite for desert dust (Lensky and Rosenfeld, 2008) makes use of the spectral absorption characteristics of aeolian desert dust. While desert dust generally is an inhomogeneous aerosol with particles form various different mineral species, size ranges and shapes, and varies in its exact properties by region (Formenti et al., 2011), its bulk radiative properties typically comprise an absorption maximum around 10 μm (Ackerman, 1997). Within the SEVIRI channels, highest dust absorption rates are reached in the 10.8 μm band, with lower absorption in the adjacent bands centred at 8.7 and 12.0 μm. The desert dust RGB composite uses brightness temperature differences (BTDs) between the high and lower absorbing bands, and assigns colours as given in Table 1. Values outside of the min–max range are clipped to the value of range boundaries. The three channels are then normalised, and minimum intensity is assigned 0, maximum intensity is assigned 1. Additionally, gamma correction is applied by exponentiating the normalised values with the factor $\frac{1}{\Gamma}$, with Γ as specified by Lensky and Rosenfeld (2008). In a last step, all three channels are combined into one RGB image.

If desert dust is present in the atmospheric column, it suppresses the BT signal at 10.8 μm more strongly than in the adjacent channels, which results in a positive signal in the red and a negative signal in the green channel. Limits of the blue channel are chosen so that it typically saturates during clear sky conditions during day, which results in dust appearing in bright pink colours (strong red, weak green, strong blue). Besides dust, this product highlights high and therefore cold cloud tops which typically appear dark red due to an absence of blue. Note that the general colour of the image is also determined by a series of other factors such as surface emissivity, surface skin temperature, the atmospheric temperature profile, presence of other aerosols and trace gases, or water vapour in the atmospheric column. Since many of these controlling factors undergo a diurnal cycle, the colour of dust changes from bright pink over hot desert surfaces during day, to washed-out purple during night, or during conditions with increased water vapour in the atmospheric column (Brindley et al., 2012). Particularly dust at altitudes below 1 km often is almost undetectable, hidden by atmospheric water vapour in the column above (Banks et al., 2019).

2.3. Feature detection

We apply automated detection methods to identify characteristic weather features in the RGB images and underlying BT data. These methods are threshold-based and do not require manual tuning or human supervision.

Firstly, we detect convective storms by their high and therefore cold cloud top in the 10.8 μm channel. This is equivalent to an absence of blue colour in the RGB image, as convective storms would often show as dark areas. We use a threshold of −40 °C (233.15 K) on the 10.8 μm channel (blue colour) to generate a binary mask of convective storms. This threshold was shown appropriate for detecting and tracking convective systems in the Sahel (Goyens et al., 2012) and West Africa (Crook et al., 2019), and is commonly used for highlighting coldest areas in visualisations of the 10.8 μm channel.

Secondly, we detect dust using the automated method from Ashpole and Washington (2012). This method identifies dust by detecting anomalously pink areas in the image, using thresholds for the three RGB channels and an anomaly threshold for deviations from a 15-day cloud-screened background value in the green channel, as given in Table 2. This detection method was shown to perform on-par with a human observer; pixels flagged as dusty agree with subjectively classified pixels at a similar fraction as subjective classifications from different human observers. The method works best when the colour of dust has a strong spectral contrast from the background colour. Limitations to the reliability of the method arise from conditions that affect this spectral contrast. This can be through changes in the colour of dust due to the height of the dust layer, the water vapour above, or the mineralogical composition of dust (see Section 2.2), but also the thermal contrast between dust and the surface. This results in dust detection being more reliable during day than during night, and more reliable for optically thick and high altitude dust during dry conditions than for thin and low altitude dust during conditions with high moisture levels. These limitations must be taken into account when interpreting results from this detection method. Our detection method for convective storms only selects coldest pixels via an absolute threshold, therefore it is not affected by a diurnal cycle in reliability.

2.4. Data preprocessing

We use SEVIRI data from the years 2018, 2019, and the first 7 months of 2024, selected based on data availability. We preprocess data as follows: We select a domain that covers West Africa (3–38 °N, 18 °W–20 °E, see Fig. 1a), from the Mediterranean to the Gulf of Guinea, and from Chad to the Atlantic. Every SEVIRI image is checked for missing values and images from timestamps that contain missing values are discarded. The remaining observations are then remapped from the native SEVIRI grid to a regular grid of 256 × 256 data points using bilinear remapping in order to simplify projection handling and reduce data resolution. This results in a mean pixel size of about 15 × 15 km. From the remapped data, BTD(120-108), BTD(108-087) are computed. In the final step, the obtained BTDs and the BT at 10.8 μm are min-max normalised to values from 0 and 1 using the limits given in Table 4. Values exceeding these boundaries are clipped to the respective min and max values. This is equivalent to the boundaries for the red and green channel in the dust RGB, but includes information from a wider value range in the 10.8 μm channel to capture growth and decay of convective systems. The data from the years 2018 and 2019 contains about 60,000 frames and is used for model training and validation. We apply a 80 % to 20 % split for training and validating the model, applied

Table 3
Summary of the data used to train, validate and test our model. Frames covers all valid timestamps with SEVIRI image data available for this study. Error filtering during preprocessing reduces this number by discarding frames with missing data. The number of frames with the pixel count of detected features exceeding 1 % of image pixels gives an estimate of the frequency of occurrence of dust and convection (cold clouds) in this dataset.

Year	Frames	Error filtering				Features exceeding 1% of image pixels			
		discarded		clean		dust		convection	
2018	35 014	76	(0.2%)	34 938	(99.8%)	11 495	(32.9%)	28 337	(81.1%)
2019	23 333	61	(0.3%)	23 272	(99.7%)	8817	(37.9%)	19 453	(83.6%)
2024[a]	16 965	403	(2.4%)	16 562	(97.6%)	9199	(55.5%)	13 365	(80.7%)

[a] Only months January through July.

Table 4
Normalisation limits for data input into DustCast. Channel information from SEVIRI Brightness Temperate (BT) or Brightness Temperature Differences (BTD).

Channel	BT_{min}(K)	BT_{max}(K)
BTD(120-108)	−4	2
BTD(108-087)	0	15
BT(108)	200	320

in blocks of 800/200 consecutive frames. This reduces autocorrelation between the training/validation sets and spreads both sets over the full year. The remaining data from the first 7 months of 2024 are used for evaluating the model on previously unseen data. These cover boreal spring and summer, which are the most active seasons for dust storms in West Africa. For a data summary see Table 3.

3. Model implementation and verification metrics

3.1. Model implementation

Fig. 1 shows a schematic overview of the data flow and sampling procedures. We implement the model using the method of Denoising Diffusion Probabilistic Models (DDPM) as introduced by Ho et al. (2020), with the diffusion process operating directly in pixel space. We implement the denoiser using a modified U-Net model (Ronneberger et al., 2015) with several layers of ResNet (He et al., 2016) blocks, spatial attention layers and 2D convolutions. Skip connections between the encoding and decoding path help with the reconstruction of small-scale features in the decoding path. An overview of the architecture is shown in Fig. A.9. For learning the denoising process, a linear beta schedule adds Gaussian noise over 1,000 iterations to all three image channels of the target frame. The denoiser is directly conditioned with the previous three frames of SEVIRI BT observations relative to the target, by directly concatenating conditional input and the noisified target. The conditioned-noisified tensor is passed into the model at top-layer, positional time step embeddings are passed into each model layer in the encoding and decoding branch. For sampling, we use the method introduced by Song et al. (2021) for Denoising Diffusion Implicit Models (DDIM), with 100 denoising steps for generating one new frame. Further consecutive frames are then generated recursively, by updating the conditional input with the newly sampled frame (Fig. 1b). Starting the sampling process from different noise fields allows the generation of multiple realisations and the creation of ensemble predictions. For generating the dust RGB images (Fig. 1c), the predicted 10.8 µm channel must be transformed back to BT before it can be normalised with the narrower limits used for the blue image channel. Data for the red and green channels are already normalised with the same limits as used for the dust RGB. Applying gamma correction for the green channel and merging the colour channels into one image generates the final RGB output.

3.2. Model training and tuning

Training the model from scratch with 256 × 256 pixels data requires considerable computational resources. For speeding up convergence, we apply an iterative training approach with pre-training at lower resolution before fine-tuning on high-resolution data. For this, we downsample the images using bilinear interpolation with antialiasing. We further run grid-based hyperparameter sweeps in order to find hyperparameter values that achieve best results with the existing model. This results in the following training procedure: We start training over 50 epochs with reduced image resolution data of 64 × 64 pixels, using the AdamW optimizer (Loshchilov and Hutter, 2019) with a MSE loss. We use an aggressive learning rate schedule (Smith and Topin, 2019) and a maximum learning rate of 2.5×10^{-5} to speed up training at early stages. In this stage, the model learns the general colour of the images as well as large-scale motion. We then retrain the model with data of 128 × 128 pixels, followed by final training with "full" 256 × 256 pixels resolution, during which the model learns small-scale features. Both retraining steps are performed over 20 epochs using a cosine scheduler. The cosine scheduler adjusts the learning rate with each cycle, starting from an initial value of 1.0×10^{-5} and gradually decreases the learning rate following a cosine curve. Further details on the training procedure can be found in Appendix A.1.

3.3. Fractions skill score for nowcast evaluation

Conventional image-generation methods often use pixel-wise evaluation metrics for assessing their performance. These metrics are not well suited for the evaluation of weather forecasting since only a small spatial or temporal displacement of features causes large penalties, even though key patterns might be reproduced. To overcome this double-penalty issue and account for good forecasts with spatial displacement, neighbourhood methods were developed. Fractions Skill Score (FSS) (Roberts and Lean, 2008) is a commonly used neighbourhood method that evaluates deterministic forecasts on different spatial scales. As a first step, binary masks (BM) are created from predicted and observed fields by applying the feature detection methods for convective storms and dust as explained in Section 2.3. Binary masks have the value 1 where a feature is detected, and 0 elsewhere. The next step computes neighbourhood probabilities (NPs) around the point k:

$$\mathrm{NP}_k = \frac{1}{L}\sum_{l=1}^{L}\mathrm{BM}_{k,l}\,, \tag{1}$$

with $\mathrm{BM}_{k,l}$ the value of the binary feature mask at the point l, and L the total number of points in the neighbourhood. NPs are the fraction of points in a neighbourhood where a feature is detected. In this study, we use square neighbourhoods centred around the point k, with the square's edge length defined as scale parameter. The scale parameter is selected based on the typical size scale of features of interest. Using NPs from prediction (NP_{prd}) and observation (NP_{obs}) fields, the Fractions Brier Score (FBS) can be computed:

$$\mathrm{FBS} = \frac{1}{K}\sum_{k=1}^{K}(\mathrm{NP}_{k,prd} - \mathrm{NP}_{k,obs})^2\,. \tag{2}$$

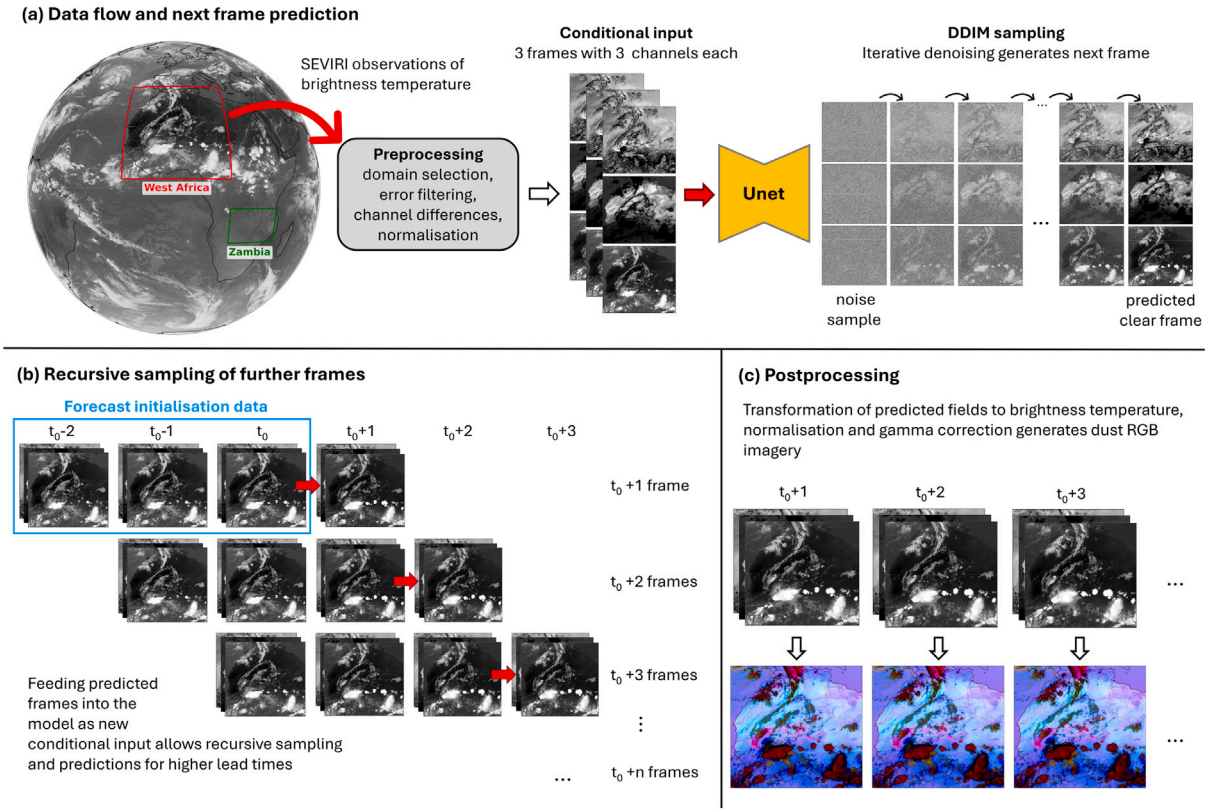

Fig. 1. Schematic overview of the domains, data flow and the sampling procedures used in this study for generating nowcasts. Note that DustCast is primarily focused on predicting the desert dust RGB composite for West Africa. DustCast uses brightness temperature observations from the same three SEVIRI channels that are used for generating this composite as conditional model input. A simplified spinout version of DustCast has been deployed for Zambia.

FBS compares fractions of predicted to fractions of observed features in all neighbourhoods within a domain. A perfect alignment yields a score of 0, larger scores indicate worse agreement between predicted and observed features. FBS strongly depends on the frequency of features. Normalisation with a reference score can be used to generalise FBS. The worst possible FBS is observed when there is no overlap between prediction and observation and can be computed as:

$$\mathrm{FBS}_{worst} = \frac{1}{K}\left[\sum_{k=1}^{K}\mathrm{NP}_{k,prd}^2 + \sum_{k=1}^{K}\mathrm{NP}_{k,obs}^2\right]. \tag{3}$$

FSS is then defined as:

$$\mathrm{FSS} = 1 - \frac{\mathrm{FBS}}{\mathrm{FBS}_{worst}}. \tag{4}$$

Perfect skill yields a FSS score of 1, no skill yields 0. Values above 0.5 are typically considered skilful (Roberts and Lean, 2008). For extending FSS to ensemble predictions, Schwartz et al. (2010) introduce a probabilistic FSS (pFSS) which replaces the prediction NPs with neighbourhood ensemble probabilities (NEPs):

$$\mathrm{NEP}_{k,prd} = \frac{1}{L}\sum_{l=1}^{L}\mathrm{EP}_{k,l}, \tag{5}$$

with $\mathrm{EP}_{k,l}$ the predicted ensemble probability at point l in a neighbourhood of point k. The equation for pFSS therefore is:

$$\mathrm{pFSS} = 1 - \frac{\sum_{k=1}^{K}(\mathrm{NEP}_{k,prd} - \mathrm{NP}_{k,obs})^2}{\sum_{k=1}^{K}\mathrm{NEP}_{k,prd}^2 + \sum_{k=1}^{K}\mathrm{NP}_{k,obs}^2}. \tag{6}$$

Scores from pFSS behave in analogy to FSS but take into account ensemble probabilities at the neighbourhood level. Necker et al. (2024) recommend pFSS over other ensemble FSS variants since it is well behaved and scales with ensemble size. In this study, we use FSS for evaluating deterministic predictions (persistence) and pFSS for evaluating ensemble predictions.

4. Results and discussion

For evaluating the performance of our model for actual nowcasting purposes, we generate a large number of predictions from observation data that were previously unseen by the model. Each nowcast is initialised with three observation frames, with the latest input time assigned as initialisation time. Using recursive sampling, we predict the following 24 frames. We maintain the time-spacing of SEVIRI data, which results in predictions in 15-minute increments. This makes a +24 frames prediction equivalent to the next 6 h. For each time step, the sampling starts from five different noise samples, generating an ensemble prediction with five members. Using this procedure, we generate nowcasts initialised at every full hour from January to July 2024, unless one of the input frames is missing or contains missing values. This results in about 3,900 nowcasts for evaluation.

4.1. Case study 7 June 2024

We present an example case for a nowcast over West Africa on 7 June 2024. Fig. 2 (top row) shows the observed dust RGB composite for this event. This particularly active weather situation contains a large number of different meteorological features that can be observed during the West African monsoon season: deep convective systems in the Sahel (dark red clusters), a large-scale convectively generated dust storm ("haboob", bright pink cluster) over Mauritania, and further features of blowing dust (pink) or high cirrus clouds (dark red line structures) scattered across the domain.

Fig. 3 shows simulated Outgoing Longwave Radiation at Top of Atmosphere (OLR) and Dust Optical Depth at 550 nm (DOD) from the Unified Model (UM) (Willett et al., 2025), currently operational at the Met Office. OLR serves as proxy for clouds, where high values indicate cloud free conditions (radiation largely from the warm surface),

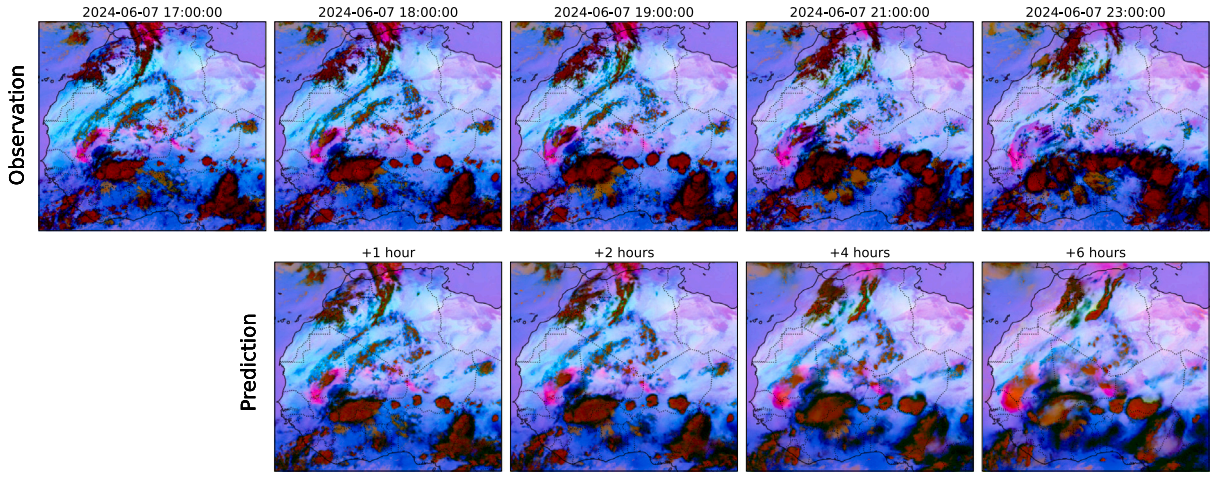

Fig. 2. Observed and predicted desert dust RGB images for a nowcast initialised at 07/06/2024 17 UTC (t_0). The top left shows the latest observation which was fed into the model and observations for t_0 +1 h, +2 h, +4 h, and +6 h. The bottom row shows the corresponding predicted frames. Dark red colours show high and cold cloud tops, bright pink colours show dust, light blue and light purple mainly show stationary background features.

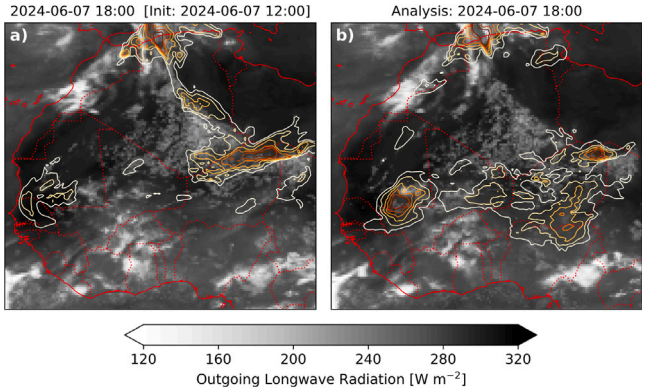

Fig. 3. Simulated outgoing longwave radiation at top of atmosphere (shading) and integrated dust optical depth at 550 nm (contours, from 0.6 to 1.2 in increments of 0.1) from the operational Met Office Unified Model; (a) forecast for 07/06/2024 18 UTC from the initialisation at 07/06/2024 12 UTC, and (b) analysis for 07/06/2024 18 UTC.

and low values indicate high clouds (lower OLR compared to surface due to their lower temperature). The UM prediction and analysis for 18 UTC both show low OLR values over southern Mali and Burkina Faso, indicating high clouds. This matches with the dark red cluster in the dust RGB in the same area. Compared to the dust RGB, the UM lacks the smaller convective cells in the Sahel (dark red clusters over Niger) in both the simulation and the analysis. For dust, the UM shows highest DOD values over northern Niger and northern Algeria in the forecast for 18 UTC, and patches of high DOD values exceeding 0.7 over Mauritania, however, without distinct structures. The dust RGB prominently shows the haboob dust storm (bright pink cluster) over the same area at 17 UTC (and after). It takes until the next model initialisation at 18 UTC that the UM simulation shows peak DOD values with a distinct front over Mauritania and Mali, extending towards the Senegal. The UM does not reproduce the haboob dust storm during the previous simulation period, and peak DOD values only appear in the model after the assimilation of observations. This demonstrates a typical situation where our nowcast can improve forecasts on short timescales and could be used to issue warnings for events that NWP does not reproduce.

Fig. 2 (bottom row) shows the first ensemble member of a RGB nowcast for the same event, initialised at 17 UTC. A visualisation of the complete ensemble nowcast can be found in appendix Fig. A.10. Detected features can be found in appendix Fig. A.11. An animation of the nowcast can be found in the supplementary material.

On visual inspection, key strengths and weaknesses of such a nowcast can be identified: the nowcast reproduces the general colour of the dust RGB and the evolution of key features such as the large haboob dust storm over Mauritania, the growth of convective storms in the Sahel area, or propagation of high clouds in the northwestern part of the domain, while surface features remain stationary. The propagation speed of the haboob and the timing that it reaches the Senegal is reproduced realistically in the nowcast. Limitations can be identified when focusing on convective activity in the Sahel. Convective features already contained in the initialisation frame grow and decay at reasonable rates, although their extent is slightly underestimated. Features that are not yet contained in the initialisation data are not reproduced in the prediction, indicating that the model struggles to reproduce convection initiation. This can be observed over Niger at lead times of 2 h and after, where new dark red features appear in the observations, but are missing in the prediction. Such failure to predict convection initiation is common across many nowcasting methods (Smith et al., 2024). Further limitations can be seen in the propagation speed of the haboob beyond 2 h lead time, which exceeds the observed speed. We hypothesise that the information contained in the input data and the limited model complexity are insufficient for reproducing the evolution of complex atmospheric flow patterns at longer lead times. The prediction further shows blurring with increasing lead time. This causes small-scale features to disappear after about 4 h and significantly changes the appearance of the image so that it limits the interpretability to a human observer.

Summarising, DustCast shows the capability of realistically predicting several frames of the dust RGB composite including key features, their propagation and development if features are already contained in the initialisation data. The limited capability of the model to produce entirely new features, uncertainties in the evolution of flow patterns, and blurring at high lead times pose limitations to the product and the maximum lead time for a useful nowcast.

4.2. Probabilistic feature-based evaluation

For creating probabilistic nowcasts, we generate ensemble nowcasts with 5 members, with the ensemble size chosen by a trade-off between sampling time and image quality. For a systematic feature-based evaluation, we automatically detect convective cells and dust in the images by applying the criteria explained in Section 2.3. We average the obtained binary feature masks over all ensemble members. This

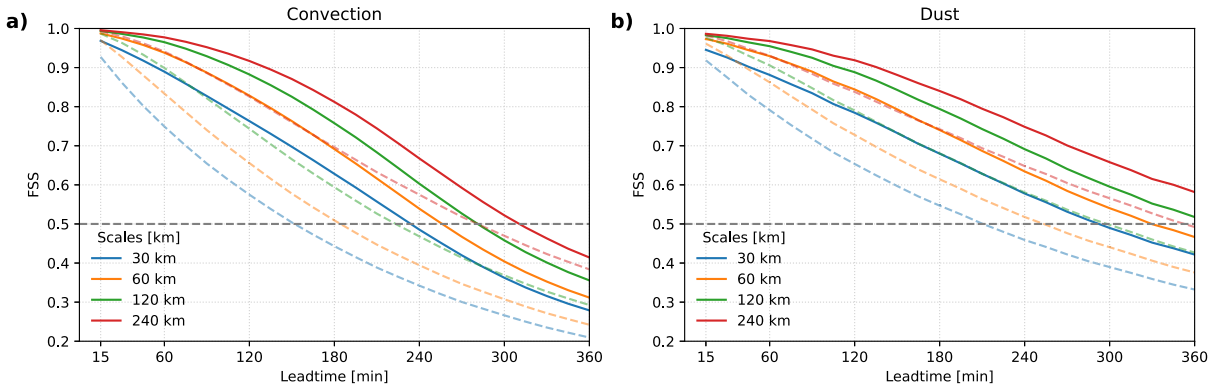

Fig. 4. Mean fractions skill scores (FSS) from about 3,900 nowcasts over lead times up to 6 h; for (a) detected convective cells, and (b) detected dust. FSS window size over scales from 30 to 240 km (2 to 16 pixels). Model predictions as solid, persistence as dashed lines. FSS of 0.5 (useful skill) as dashed grey reference line.

generates probabilistic fields for dust and convective cells. To avoid edge-effects at the domain boundaries, we introduce an 8 pixels crop to the domain edges. The evaluation therefore covers a domain of about 4–37 °N, -17–19 °E. To minimise the issues introduced through the limitations of the dust detection method at night, we only evaluate dust for forecast validity times between 06 and 18 UTC. We further exclude timestamps from the respective analysis where less than 1% of observed pixels are detected as convective cells (cold clouds) or dust. This minimises misleading scores from events where the scene is largely free from convection or dust.

Fig. 4 shows mean pFSS for convection and dust for the ensemble nowcast and the corresponding FSS for a persistence forecast, evaluated at spatial scales of 30 km, 60 km, 120 km, and 240 km. While FSS on small scales shows the fastest decline with lead time, our model consistently outperforms the persistence forecast, and achieves useful skill (FSS > 0.5) for convection up to about 4 h lead time, and for dust for up to 5 h on all analysed scales. In general, skill on small scales declines faster with increasing lead time than on large scales. This fits with processes on the size-range of individual convective storms operating on much shorter timescales than meso- and synoptic-scale systems and processes. Any future observation therefore shows more differences on small scales than on large scales when compared to initialisation, and requires the nowcast to reproduce more small-scale features to achieve good scores. Note that the persistence forecast also achieves scores of useful skill for lead times of multiple hours. The skill benefit compared with persistence increases with lead time, and is greater for smaller scales. Besides typical timescales of processes at these scales, this can be related to increasing blurring of the model nowcast, leading to a reduction of total feature points at large lead times. Comparing lead time for forecasts with useful skill, our nowcast on average extends the maximum lead time by about 1.5 h for convection, and about 1.25 h for dust compared to a persistence forecast and for features operating on 30 km scales, reducing to about 45 min for convection and 1 h for dust on 240 km scales.

For evaluating probabilistic forecasts, receiver operating characteristic (ROC) curves assess how well predicted events correspond to observed occurrences and non-occurrences of events across varying probability thresholds. For each probability threshold, the probabilistic nowcast fields are converted to binary fields with value 1 if the predicted probability is greater equal the threshold value, and 0 elsewhere. True positive rates (forecast and observation both 1, relative to all observed points with 1) are then compared to the false positive rate (forecast 1 where observation 0, relative to all observed points with 0). A perfect forecast maximises the true positive rate while minimising the false positive rate. If true positive rates are plotted against false positive rates, a perfect forecast would result in a curve in the top left corner of such diagram. The area under the curve (AUC) is commonly used as a binary metric for expressing probabilistic forecast skill based

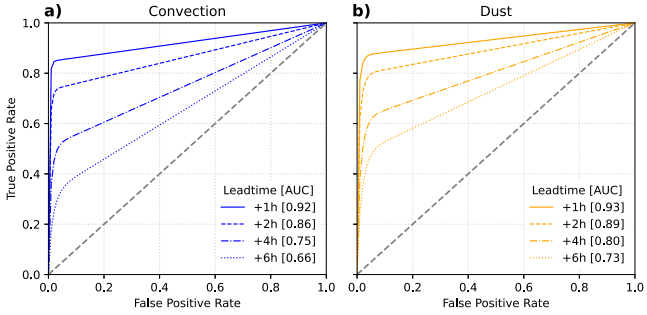

Fig. 5. Mean receiver operating characteristic (ROC) curves at lead times of 1, 2, 4, and 6 h; for (a) detected convective cells, and (b) detected dust. Performance of a random classifier (no skill) as dashed grey reference line.

on ROC curves. A perfect forecast achieves an AOC of 1, a random classifier achieves AUC equals 0.5. Fig. 5 shows mean ROC curves for convection and dust for different lead times. All curves remain above the reference level of a random classifier (diagonal line), with the AUC achieving very high scores of above 0.9 at lead times of 1 h for both convection and dust. Note the high values of true positive rates for very low false positive rates. This indicates the model correctly forecasting most events while producing only a small number of false alarms.

Fig. 6 shows AUC for convection and dust over the time of day for which the nowcast is valid and for different lead times. The AUC curve for convection shows highest values for nowcasts valid between morning and noon and lowest values for the afternoon. This is consistent between nowcasts with different lead times, with a positive time-offset in highest (lowest) values with increasing lead time and aligns with the diurnal cycle of convection initiation, which is lowest in the morning and peaks in the afternoon. This confirms previous observations that the model reproduces growth and decay of convective systems, but struggles with convection initiation after model initialisation. For dust, the AUC curves show a less consistent diurnal cycle in skill between different forecast lead times compared to convection. Note that dust detection is significantly affected by uncertainty introduced through the dust detection method as explained in Section 2.3. In fact, the colour change of the dust RGB in the morning and evening results in very different feature masks when comparing forecasts initialised at night/day to observations from day/night. While the two key processes leading to dust-emitting peak-wind-speeds in this domain show a characteristic diurnal cycle, with the down-mixing of the nocturnal low level jet (Knippertz, 2008) causing gusts in the morning, and cold pools from convective systems leading to peak-wind-speeds at night (Marsham et al., 2011), the nowcast skill cycle for predicting dust cannot reliably be attributed to those.

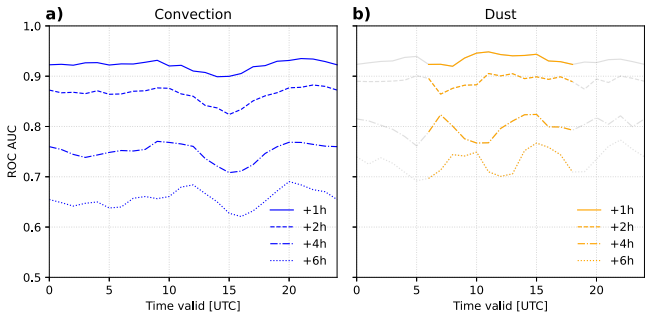

Fig. 6. Mean receiver operating characteristic (ROC) area under curve (AUC) score over time of day that the nowcast is valid, at lead times of 1, 2, 4, and 6 h; for (a) for detected convective cells, and (b) detected dust. Note that nighttime dust before 6 UTC and after 18 UTC is excluded in our analysis due to the double-penalty introduced by limitations of the dust detection method during night.

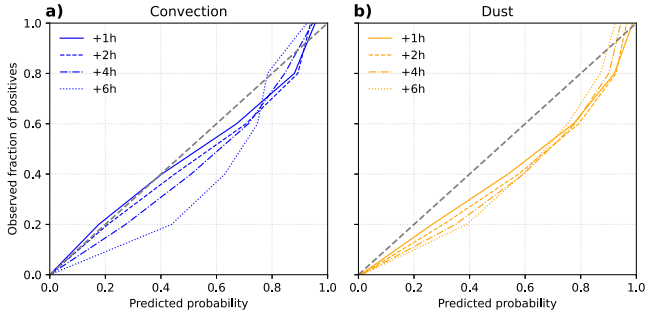

Fig. 7. Mean calibration curves at lead times of 1, 2, 4, and 6 h; for (a) detected convective cells, and (b) detected dust. Perfect calibration as dashed grey reference line.

Calibration curves assess how well predicted probabilities align with the observed relative frequency of cases. A perfect forecast predicts probabilities that match how often events are observed, which results in a calibration curve that lies on the diagonal. Fig. 7 shows mean calibration curves for convection and dust at different lead times. For convection, the curves closely follow the diagonal line of a perfectly calibrated prediction for lead times of up to two hours and predicted probabilities up to about 0.5, and slightly fall under the diagonal for higher predicted probabilities. This means the model predicts high probabilities for convection being present more frequently than it is observed, which is known as overconfidence. For increasing lead times, the curve falls lower under the diagonal, indicating increasing overconfidence for predicting convection. For dust, all curves are located closely together but below the diagonal. This indicates the model generally being overconfident for predicting the presence of dust (or underconfident for predicting clear conditions). Nevertheless, the calibration curve for dust predictions roughly follows the diagonal trend of a well-calibrated model.

4.3. Computational requirements

Diffusion models are computationally more expensive than conventional machine learning methods since they require multiple passes through the model to generate one output sample. To allow DustCast being run in rapid-refresh cycles, we aim for a +6 h ensemble nowcast to be generated in less than 10 min on one of the GPUs used for training. This allows for the complete nowcast to be available before the next SEVIRI observations are received. For our model, one frame of the ensemble ($5 \times 256 \times 256$) requires about 19 s of computation time, which results in about 7.5 min for generating a full +6 h ensemble nowcast. We tested larger ensemble sizes with fewer sampling steps that maintain the total sampling time. However, the reduced number of sampling steps lead to instabilities in the recursive sampling process, worsened the model calibration and did not significantly improve pFSS (not shown). We suggest that future developments make use of improved samplers that can be run with a lower number of sampling steps. The reduced sampling time could be used to reduce the overall latency for producing the nowcast, or to increase the number of high-quality ensemble members. Alternatively, the prediction process could be run in parallel on multiple GPUs for increasing the number of ensemble members while maintaining latency. Note that our implementation of DustCast is a rather simplistic one, and that computational requirements could likely be reduced with a latent-space diffusion model.

4.4. Observations from live deployment

We deployed a spin-out version of DustCast during a nowcasting testbed within the Weather and Climate Information Services (WISER) - Early Warnings for Southern Africa (EWSA) project in Zambia in February 2025. The testbed focused on researchers working with local forecasters to use a variety of nowcasting tools to improve predictions of severe storms. For this application, we deployed a simplified model that was trained with data from southern Africa and only predicts the SEVIRI 10.8 µm channel. This provided real-world testing of DustCast in Africa, and new insights into DustCast's performance for rapidly developing convective storms. We therefore report several key observations from this live deployment:

Firstly, conventional OF-based nowcast products showed limited capability to reproduce cases of rapid storm growth and intensification. In these situations, DustCast clearly outperformed these conventional methods by successfully predicting storm growth and decay. Fig. 8 shows an example case of a rapidly growing and intensifying convective system in the vicinity of the testbed focus area (red cross, growing system west from marker). Initiation and intensification happened over a short period of only two hours. As soon as initial storm growth (sinking BTs) were contained in the model input data (initialisation at 13:30 UTC), the nowcast realistically predicted further intensification of the system. In comparison, nowcast systems based on OF predicted low rain rates but missed the intensification (appendix Fig. A.12). In such cases, DustCast can help forecasters to identify expected intensification that is missed in OF-based nowcasts. Local observers reported 31 mm of rain in the focus area, accumulated from 14:30 and 16:00 UTC. This resulted in flooding of informal settlements, directly impacting people's daily activities and livelihoods. Taking into account the latency of receiving satellite data and running DustCast, our nowcast could have been used for issuing warnings 30 min ahead of the flooding. Such early warning is crucial for mitigating adverse impacts, and highlights the potential for operational deployment of DustCast. Secondly, predicting convection initiation remains challenging for both OF-based products and DustCast due to its inherent stochastic nature. Thirdly, the accessibility of the product is a significant hurdle to operational forecasters making use of the nowcast. A nowcast of an existing visualisation such as false-colour RGB imagery or the 10.8 µm cloud channel provides significant advantages over entirely new products, since forecasters are already trained in their interpretation. This reduces the hurdle to adoption and ultimately allows improvements of forecasters' predictions and early warnings issued.

5. Conclusions

We introduce DustCast, a simple diffusion model for generating forward-predictions of the existing desert dust RGB composite. We train the model on two years of SEVIRI observations from a large domain over West Africa and show that our nowcasts can reproduce

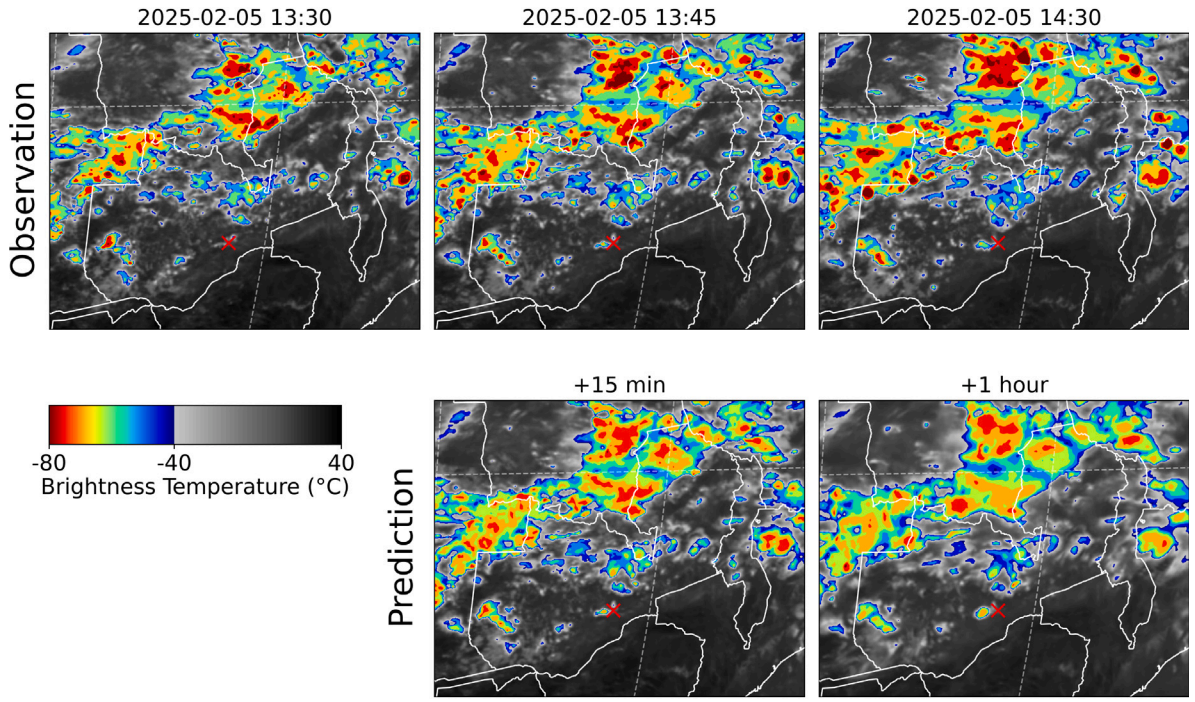

Fig. 8. Observed and predicted SEVIRI 10.8 μm channel for a nowcast over Zambia, initialised at 05/02/2025 13:30 UTC (t_0). The top left shows the latest observation which was fed into the model. The remaining top row plots show observations for t_0 +15 min and +1 h. The bottom row shows the corresponding predicted frames. Lusaka (main focus area of testbed) is marked with a red cross. Temperatures lower than −40 °C are highlighted. Note this nowcast is from a testbed-specific spin-out version of DustCast which only predicts the 10.8 μm channel, and that values colder than 200 K are clipped in the model.

convectively generated dust storms. Our method offers advantages compared to conventional OF-based nowcasts by predicting complete images of complex weather situations. This includes other features besides dust such as convective systems, and helps with assessing weather context. Compared to previous nowcasts of other RGB composites (Ionescu et al., 2021; Bari et al., 2023), our implementation using a diffusion model creates sharper realisations with the option to generate ensemble nowcasts. Combined with feature detection within the predicted frames, the ensemble allows a probabilistic nowcast and scenario assessment. We generate probabilistic nowcasts for dust and convective (cold cloud) features, which show skill at lead times of 5 h, and 4 h, respectively, on scales of 30 km. On large scales of 240 km, the dust nowcast shows skill at 6 h (and more), the convection nowcast shows skill at 5 h lead time. In comparison to previous OF-based nowcasts of SEVIRI-derived precipitation (Burton et al., 2022), our nowcast extends skill by more than two hours. However, note that feature detection introduces additional uncertainty through the reliability of the detection method. Since dust detection from infrared channels alone remains challenging, the visualisation as RGB composite still holds value and leaves the interpretation to a human observer.

While DustCast performs well for predicting the growth and decay of features, the prediction of entirely new features that are not visibly contained in the conditional input remains challenging. This particularly applies to the initialisation of convection, which is inherently unpredictable and might require advanced probabilistic methods. We observed during deployment in a testbed that another hurdle to the successful adoption of new nowcast products is their accessibility. A nowcast of an existing visualisation such as the dust RGB therefore provides an easy-to-use product for operational forecasters who are already trained in the interpretation of the RGB imagery.

Our approach could be adapted to nowcast other commonly used RGB composites, e.g., convection, ash, or fog. Our model could further be retrained with data from the new Flexible Combined Imager (FCI) on Meteosat Third Generation (MTG) satellites (Holmlund et al., 2021), which provides measurements on similar bands as SEVIRI. We suggest that future implementations make use of data transformations into lower-dimensional latent space as introduced by Rombach et al. (2022) in order to reduce computational requirements and therefore latency. Further improvements could likely be achieved by implementing a separate forecaster stack and 3D convolutions as successfully demonstrated for nowcasting precipitation (Leinonen et al., 2023). Finally, the inclusion of additional input variables such as solar angle, time of day, or orography could help to improve the colour cycle and potentially reproduce convection initiation. In the wider context, nowcasts of satellite observations can not only be used to predict other RGB composites, but indeed allow predictions of retrievals which are based on these products.

CRediT authorship contribution statement

Kilian Hermes: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **John H. Marsham:** Writing – review & editing, Supervision, Conceptualization. **Massimo Bollasina:** Writing – review & editing, Supervision. **Melissa Brooks:** Writing – review & editing, Supervision, Data curation. **Martina Klose:** Writing – review & editing, Supervision. **Franco Marenco:** Writing – review & editing, Supervision.

Code availability

All code for model training and evaluation, and for generating nowcasts is available on GitHub:
https://github.com/kfhermes/DustCast
The exact code version used for this paper is archived on Zenodo:
https://doi.org/10.5281/zenodo.17663274
The pre-trained model weights can be obtained from Zenodo:
https://doi.org/10.5281/zenodo.17508992

Table A.5
Overview of schedulers and hyperparameters tested during the grid search approach.

LR schedule	Beta schedule	LR (max)	Batch size	Epochs
OneCycleLR	linear	5.0×10^{-4}	32	50
CosAnnealingLR	cosine	2.5×10^{-4}	16	20
		1.0×10^{-4}	8	10
		1.0×10^{-5}		

Table A.6
Summary of schedulers and hyperparameters used for the final iterative training of DustCast.

Image resolution	LR schedule	Beta schedule	LR (max)	Batch size	Epochs
64 × 64	OneCycle	linear	2.5×10^{-4}	16	50
128 × 128[a]	CosineAnnealing	linear	1.0×10^{-4}	16	20
256 × 256[b]	CosineAnnealing	linear	1.0×10^{-4}	16	20

[a] Retrained from 64 × 64.
[b] Retrained from 128 × 128.

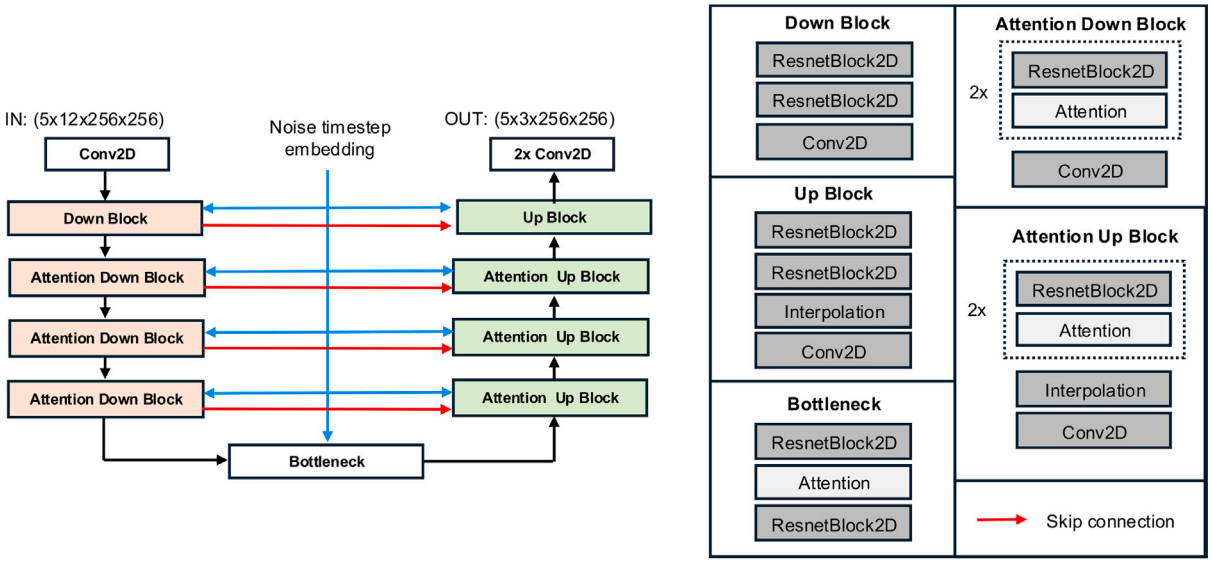

Fig. A.9. Overview of the model architecture. This is based on the UNet2DModel from the Huggingface Diffusers Python package.

Funding

KH is funded by NERC, United Kingdom through the SENSE CDT (NE/T00939X/1), is co-funded by the Met Office through a CASE studentship and received travel funding from the WISER EWSA project (W3_GRT22_SOUTH). WISER EWSA is funded by the Met Office as part of the WISER programme on behalf the UK Government's Foreign, Commonwealth and Development Office (FCDO). JM is also funded by the NERC NAIAR project (NE/Y000420/1), WISER EWSA (W3_GRT22_SOUTH) and UKRI EW4Energy (project ref. 474). MK received funding by the Helmholtz Association's Initiative and Networking Fund (grant no. VH-NG-1533).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank EUMETSAT for providing SEVIRI data. We further acknowledge package developers and maintainers, particularly for PyTorch. We thank projects and funders, particularly WISER EWSA, NAIAR, EW4Energy, and the NERC-funded SENSE CDT. KH acknowledges the latter for also providing technical training. We finally thank two anonymous reviewers for their constructive feedback that helped improving the clarity of this manuscript.

Appendix A. Model details and extra figures

A.1. Details on training and hyperparameter tuning

We used a grid-search approach over the schedulers and parameters listed in Table A.5. We trained the model using the AdamW optimizer with MSE as loss function. Data was noisified over 1,000 steps. Model performance for each training run was evaluated using mean error, root mean square error (RMSE) and structural similarity measure (SSIM). Training was performed on 4 GPUs using PyTorch Lightning in DistributedDataParallel (DDP) mode. In DDP mode, Lightning automatically shards each training batch across multiple GPUs but accumulates the gradients so that the effective batch size remains constant and independent from the number of GPUs. No further gradient accumulation was used. Due to computational constraints, training was started with downsampled images of 64 × 64 px. Downsampling was performed using the PyTorch function "Resize", with bilinear interpolation and antialiasing. The best performing model was then retrained on the same training dataset but with increased resolution. The resulting iterative training approach is summarised in Table A.6

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.wace.2025.100828.

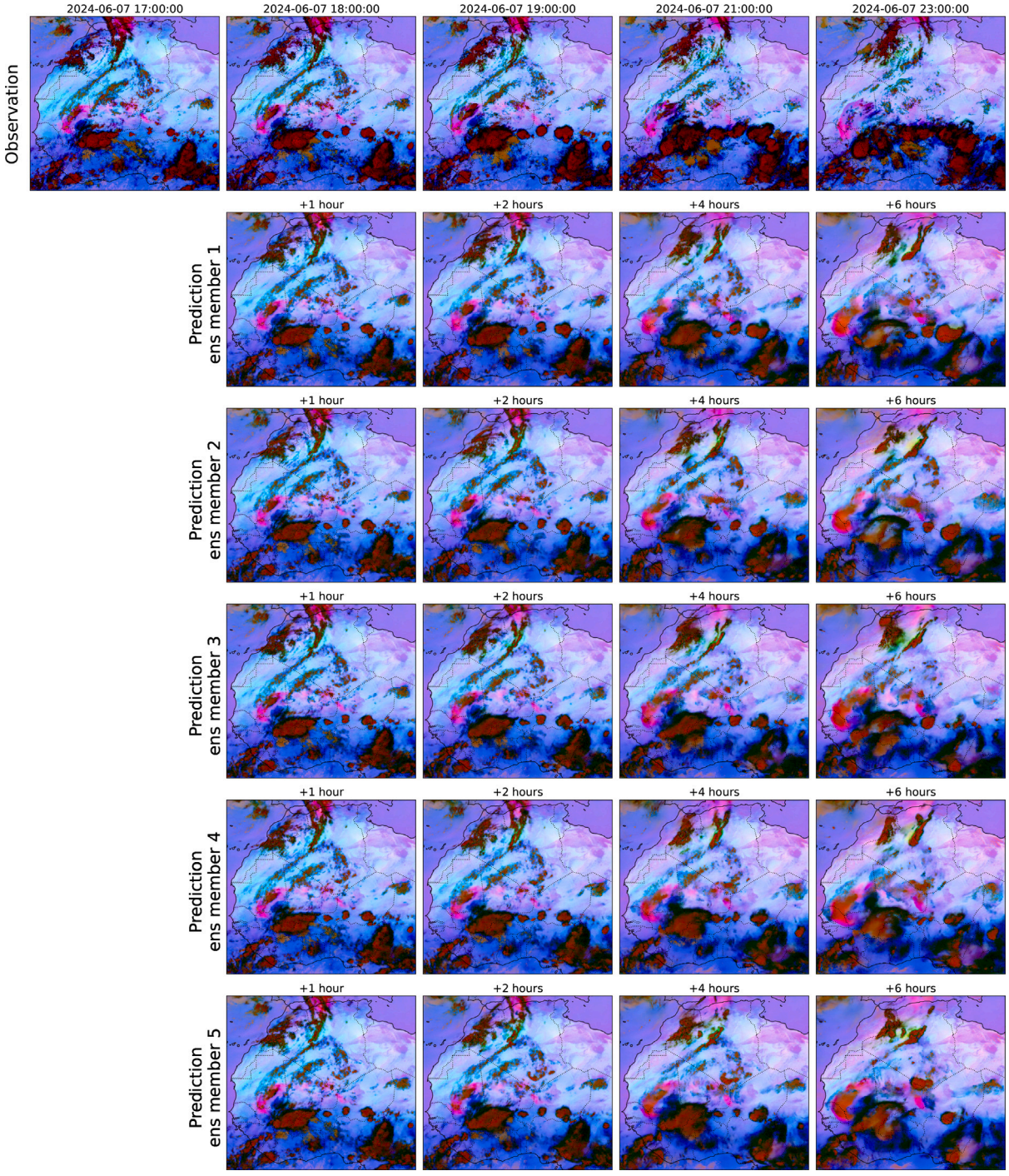

Fig. A.10. Observed and predicted desert dust RGB images for a nowcast initialised at 07/06/2024 17 UTC (t_0). The top left shows the latest observation which was fed into the model and observations for t_0 +1 h, +2 h, +4 h, and +6 h. The remaining rows show the corresponding predictions from each ensemble member. Dark red colours show high and cold cloud tops, bright pink colours show dust, light blue and light purple mainly show stationary background features.

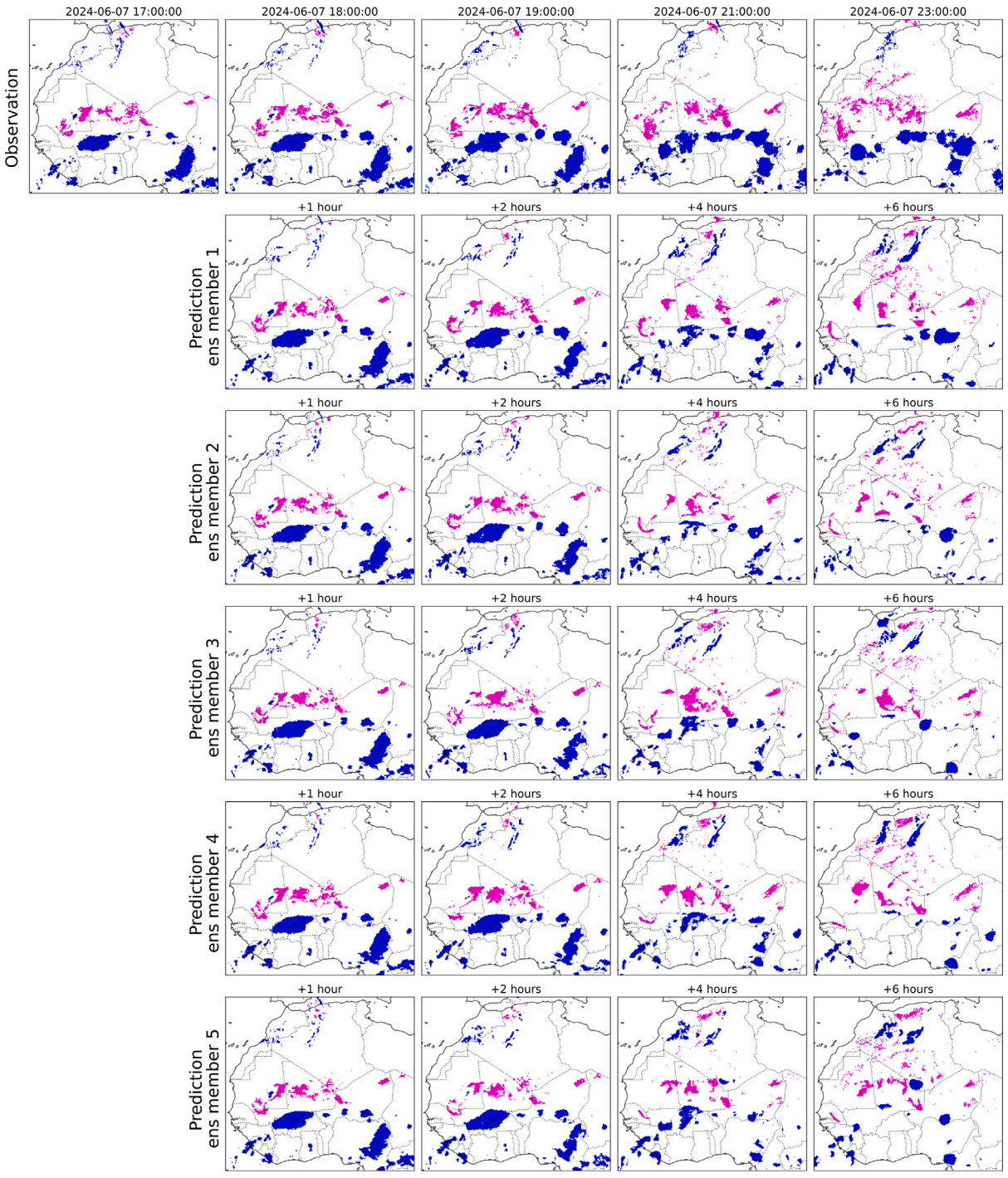

Fig. A.11. As Fig. A.10 but for dust (pink) and convection (blue) as detected in the observed and predicted desert dust RGB images.

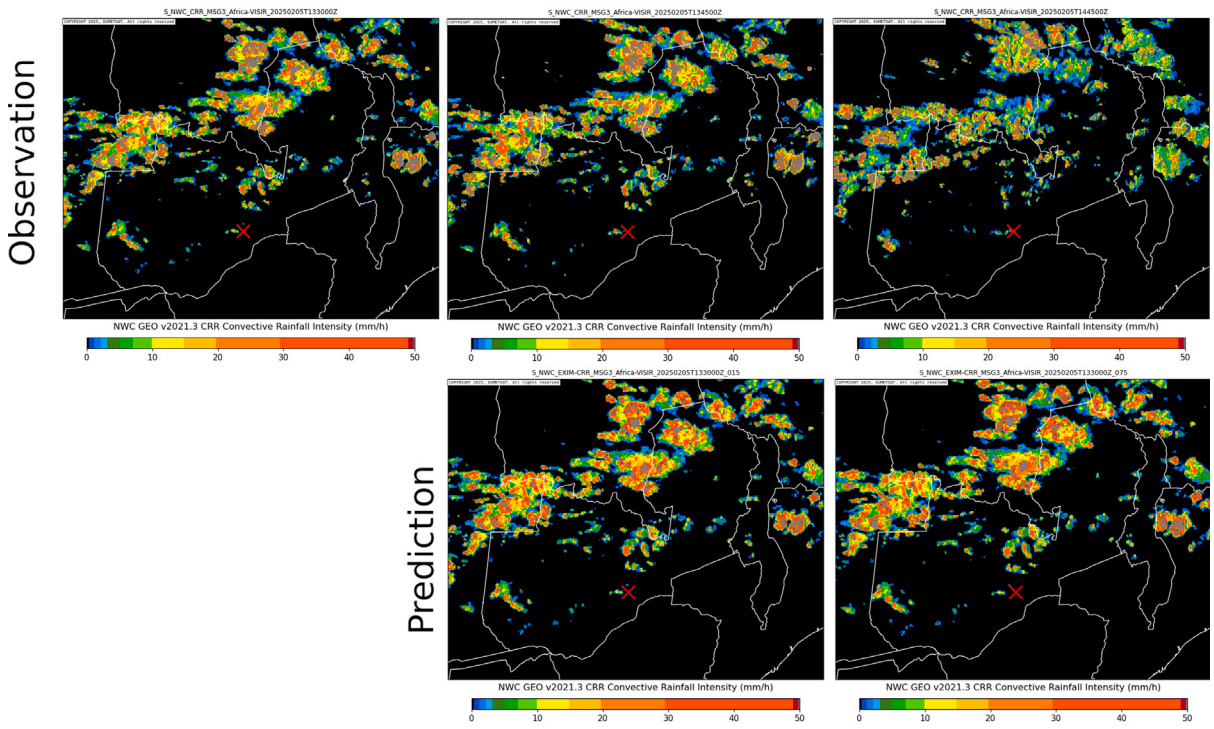

Fig. A.12. Retrieval and nowcast of Convective Rain Rate (CRR) over Zambia from the Nowcasting and Very Short Range Forecasting Eumetsat Satellite Application Facility (NWC SAF) (NWC SAF, 2013) as deployed during the WISER-EWSA testbed. This nowcast is based on optical flow and a deterministic forward propagation of the CRR retrieval. The nowcast was initialised at 05/02/2025 13:30 UTC (t_0). The top left shows the latest observation from which the forward extrapolation was performed. The remaining top row plots show observations for t_0 +15 min and +1:15 h. The bottom row shows the corresponding predicted frames. Lusaka (main focus area of testbed) is marked with a red cross. Note that the CRR retrieval for 14:30 UTC is unavailable, therefore we show data for 14:45 UTC.

Data availablity

Meteosat SEVIRI data are available from EUMETSAT: https://navigator.eumetsat.int/product/EO:EUM:DAT:MSG:HRSEVIRI.

References

Ackerman, S.A., 1997. Remote sensing aerosols using satellite infrared observations. J. Geophys. Res.: Atmos. 102, 17069–17079. http://dx.doi.org/10.1029/96JD03066.

Al-Hemoud, A., Al-Sudairawi, M., Neelamanai, S., Naseeb, A., Behbehani, W., 2017. Socioeconomic effect of dust storms in Kuwait. Arab. J. Geosci. 10, 18. http://dx.doi.org/10.1007/s12517-016-2816-9.

Alfarano, A., Maiano, L., Papa, L., Amerini, I., 2024. Estimating optical flow: A comprehensive review of the state of the art. Comput. Vis. Image Understanding 249, 104160. http://dx.doi.org/10.1016/j.cviu.2024.104160.

Ashpole, I., Washington, R., 2012. An automated dust detection using SEVIRI: A multiyear climatology of summertime dustiness in the central and western Sahara. J. Geophys. Res.: Atmospheres 117, http://dx.doi.org/10.1029/2011JD016845.

Ashpole, I., Washington, R., 2013. A new high-resolution central and western Saharan summertime dust source map from automated satellite dust plume tracking. J. Geophys. Res.: Atmos. 118, 6981–6995. http://dx.doi.org/10.1002/jgrd.50554.

Baddock, M.C., Strong, C.L., Murray, P.S., McTainsh, G.H., 2013. Aeolian dust as a transport hazard. Atmos. Environ. 71, 7–14. http://dx.doi.org/10.1016/j.atmosenv.2013.01.042.

Banks, J.R., Hünerbein, A., Heinold, B., Brindley, H.E., Deneke, H., Schepanski, K., 2019. The sensitivity of the colour of dust in MSG-SEVIRI Desert Dust infrared composite imagery to surface and atmospheric conditions. Atmos. Chem. Phys. 19, 6893–6911. http://dx.doi.org/10.5194/acp-19-6893-2019.

Bari, D., Lasri, N., Souri, R., Lguensat, R., 2023. Machine Learning for Fog-and-Low-Stratus Nowcasting from Meteosat SEVIRI Satellite Images. Atmosphere 14 (953), http://dx.doi.org/10.3390/atmos14060953.

Bhattachan, A., Okin, G.S., Zhang, J., Vimal, S., Lettenmaier, D.P., 2019. Characterizing the Role of Wind and Dust in Traffic Accidents in California. GeoHealth 3, 328–336. http://dx.doi.org/10.1029/2019GH000212.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., Tian, Q., 2022. Pangu-Weather: A 3D High-Resolution Model for Fast and accurate global weather forecast. http://dx.doi.org/10.48550/arXiv.2211.02556.

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

Brindley, H., Knippertz, P., Ryder, C., Ashpole, I., 2012. A critical evaluation of the ability of the Spinning Enhanced Visible and Infrared Imager (SEVIRI) thermal infrared red-green-blue rendering to identify dust events: Theoretical analysis. J. Geophys. Res.: Atmos. 117. http://dx.doi.org/10.1029/2011JD017326.

Browning, K.A., Collier, C.G., 1989. Nowcasting of precipitation systems. Rev. Geophys. 27, 345–370. http://dx.doi.org/10.1029/RG027i003p00345.

Burton, R.R., Blyth, A.M., Cui, Z., Groves, J., Lamptey, B.L., Fletcher, J.K., Marsham, J.H., Parker, D.J., Roberts, A., 2022. Satellite-Based Nowcasting of West African Mesoscale Storms has Skill at up to 4-h Lead Time. Weather. Forecast. 37, 445–455. http://dx.doi.org/10.1175/WAF-D-21-0051.1.

Chen, H., Zhong, X., Zhai, Q., Li, X., Chan, Y.W., Chan, P.W., Huang, Y., Li, H., Shi, X., 2025. Skillful Nowcasting of Convective Clouds With a cascade diffusion model. http://dx.doi.org/10.48550/arXiv.2502.10957.

Cowie, S.M., Knippertz, P., Marsham, J.H., 2014. A climatology of dust emission events from northern Africa using long-term surface observations. Atmos. Chem. Phys. 14, 8579–8597. http://dx.doi.org/10.5194/acp-14-8579-2014.

Crook, J., Klein, C., Folwell, S., Taylor, C.M., Parker, D.J., Stratton, R., Stein, T., 2019. Assessment of the representation of West African storm lifecycles in convection-permitting simulations. Earth Space Sci. 6, 818–835. http://dx.doi.org/10.1029/2018EA000491.

de Longueville, P., Doumbia, S., Henry, S., 2013. Desert dust impacts on human health: An alarming worldwide reality and a need for studies in West Africa. Int. J. Biometeorol. 57, 1–19. http://dx.doi.org/10.1007/s00484-012-0541-y.

Denton, E., Fergus, R., 2018. Stochastic Video Generation with a learned prior. In: Proceedings of the 35th International Conference on Machine Learning. PMLR, pp. 1174–1183.

Dhariwal, P., Nichol, A., 2021. Diffusion Models Beat GANs on image synthesis. In: 35th Conference on Neural Information Processing Systems. NeurIPS 2021, http://dx.doi.org/10.48550/arXiv.2105.05233.

EUMETSAT, 2015. MSG Meteorological Products Extraction Facility Algorithm Specification Document. Algorithm Specification Document EUM/MSG/SPE/022. v7B e-signed.

EUMETSAT, 2017. MSG Level 1.5 Image Data Format Description. Data Format Description EUM/MSG/ICD/105. v8 e-signed.

Flamant, C., Chaboureau, J.P., Parker, D.J., Taylor, C.M., Cammas, J.P., Bock, O., Timouk, F., Pelon, J., 2007. Airborne observations of the impact of a convective

system on the planetary boundary layer thermodynamics and aerosol distribution in the inter-tropical discontinuity region of the West African Monsoon. Q. J. R. Meteorol. Soc. 133, 1175–1189. http://dx.doi.org/10.1002/qj.97.

Formenti, P., Schütz, L., Balkanski, Y., Desboeufs, K., Ebert, M., Kandler, K., Petzold, A., Scheuvens, D., Weinbruch, S., Zhang, D., 2011. Recent progress in understanding physical and chemical properties of African and Asian mineral dust. Atmos. Chem. Phys. 11, 8231–8256. http://dx.doi.org/10.5194/acp-11-8231-2011.

Goyens, C., Lauwaet, D., Schröder, M., Demuzere, M., Van Lipzig, N.P.M., 2012. Tracking mesoscale convective systems in the Sahel: Relation between cloud parameters and precipitation. Int. J. Climatol. 32, 1921–1934. http://dx.doi.org/10.1002/joc.2407.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778. http://dx.doi.org/10.48550/arXiv.1512.03385.

Heinold, B., Knippertz, P., Marsham, J.H., Fiedler, S., Dixon, N.S., Schepanski, K., Laurent, B., Tegen, I., 2013. The role of deep convection and nocturnal low-level jets for dust emission in summertime West Africa: Estimates from convection-permitting simulations. J. Geophys. Res.: Atmos. 118, 4385–4400. http://dx.doi.org/10.1002/jgrd.50402.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 6840–6851. http://dx.doi.org/10.48550/arXiv.2006.11239.

Holmlund, K., Grandell, J., Schmetz, J., Stuhlmann, R., Bojkov, B., Munro, R., Lekouara, M., Coppens, D., Viticchie, B., August, T., Theodore, B., Watts, P., Dobber, M., Fowler, G., Bojinski, S., Schmid, A., Salonen, K., Tjemkes, S., Aminou, D., Blythe, P., 2021. Meteosat Third Generation (MTG): Continuation and Innovation of Observations from Geostationary Orbit. Bull. Am. Meteorol. Soc. 102, E990–E1015. http://dx.doi.org/10.1175/BAMS-D-19-0304.1.

Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T., 2017. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2462–2470.

Ionescu, V.S., Czibula, G., Mihuleţ, E., 2021. DeePS at: A deep learning model for prediction of satellite images for nowcasting purposes. Procedia Comput. Sci. 192, 622–631. http://dx.doi.org/10.1016/j.procs.2021.08.064.

Keane, R.J., Parker, D.J., Dunn-Sigouin, E.W., Marsham, J.H., 2025. Mid-Latitude Versus Tropical Scales of Predictability and Their Implications for Forecasting. Meteorol. Appl. 32, e70055. http://dx.doi.org/10.1002/met.70055.

Keisler, R., 2022. Forecasting global weather with graph neural networks. http://dx.doi.org/10.48550/arXiv.2202.07575.

Knippertz, P., 2008. Dust emissions in the West African heat trough—The role of the diurnal cycle and of extratropical disturbances. Meteorol. Z. 17, 553–563. http://dx.doi.org/10.1127/0941-2948/2008/0315.

Knippertz, P., Deutscher, C., Kandler, K., Müller, T., Schulz, O., Schütz, L., 2007. Dust mobilization due to density currents in the Atlas region: Observations from the Saharan Mineral Dust Experiment 2006 field campaign. J. Geophys. Res.: Atmos. 112. http://dx.doi.org/10.1029/2007JD008774.

Knippertz, P., Todd, M.C., 2010. The central west Saharan dust hot spot and its relation to African easterly waves and extratropical disturbances. J. Geophys. Res.: Atmos. 115. http://dx.doi.org/10.1029/2009JD012819.

Knippertz, P., Todd, M.C., 2012. Mineral dust aerosols over the Sahara: Meteorological controls on emission and transport and implications for modeling. Rev. Geophys. 50. http://dx.doi.org/10.1029/2011RG000362.

Lam, R., Sanchez-Gonzalez, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., Battaglia, P., 2023. Learning skillful medium-range global weather forecasting. Science 382, 1416–1421. http://dx.doi.org/10.1126/science.adi2336.

Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., Clare, M.C.A., Lessig, C., Maier-Gerber, L., Bouallègue, Z.B., Nemesio, A.P., Dueben, P.D., Brown, A., Pappenberger, F., Rabier, F., 2024. AIFS – ECMWF's data-driven forecasting system. http://dx.doi.org/10.48550/arXiv.2406.01465.

Leinonen, J., Hamann, U., Nerini, D., Germann, U., Franch, G., 2023. Latent diffusion models for generative precipitation nowcasting with accurate uncertainty quantification.

Lensky, I.M., Rosenfeld, D., 2008. Clouds-Aerosols-Precipitation Satellite Analysis Tool (CAPSAT). Atmos. Chem. Phys. 8, 6739–6753. http://dx.doi.org/10.5194/acp-8-6739-2008.

Loshchilov, I., Hutter, F., 2019. Decoupled weight decay regularization. In: International Conference on Learning Representations. http://dx.doi.org/10.48550/arXiv.1711.05101.

Marsham, J.H., Knippertz, P., Dixon, N.S., Parker, D.J., Lister, G.M.S., 2011. The importance of the representation of deep convection for modeled dust-generating winds over West Africa during summer. Geophys. Res. Lett. 38. http://dx.doi.org/10.1029/2011GL048368.

Middleton, N.J., 2017. Desert dust hazards: A global review. Aeolian Res. 24, 53–63. http://dx.doi.org/10.1016/j.aeolia.2016.12.001.

Necker, T., Wolfgruber, L., Kugler, L., Weissmann, M., Dorninger, M., Serafin, S., 2024. The fractions skill score for ensemble forecast verification. Q. J. R. Meteorol. Soc. 150, 4457–4477. http://dx.doi.org/10.1002/qj.4824.

NWC SAF, 2013. Algorithm theoretical basis document for convective rainfall rate (CRR-pge05 v4.0). Algorithm Theoretical Basis Document SAF-NWC-CDOP2-INM-SCI-ATBD-05_v4.0.

Roberts, A.J., Fletcher, J.K., Groves, J., Marsham, J.H., Parker, D.J., Blyth, A.M., Adefisan, E.A., Ajayi, V.O., Barrette, R., de Coning, 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., Woodhams, B.J., 2022. Nowcasting for Africa: Advances, potential and value. Weather 77, 250–256. http://dx.doi.org/10.1002/wea.3936.

Roberts, A.J., Knippertz, P. and, 2014. The formation of a large summertime Saharan dust plume: Convective and synoptic-scale analysis. J. Geophys. Res.: Atmos. 119, 1766–1785. http://dx.doi.org/10.1002/2013JD020667.

Roberts, N.M., Lean, H.W., 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Mon. Weather Rev. 136, 78–97. http://dx.doi.org/10.1175/2007MWR2123.1.

Roberts, A.J., Marsham, J.H., Knippertz, P., Parker, D.J., Bart, M., Garcia-Carreras, M., McQuaid, J.B., Rosenberg, P.D., Walker, D., 2017. New Saharan wind observations reveal substantial biases in analysed dust-generating winds. Atmos. Sci. Lett. 18, 366–372. http://dx.doi.org/10.1002/asl.765.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10684–10695. http://dx.doi.org/10.48550/arXiv.2112.10752.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer, pp. 234–241. http://dx.doi.org/10.1007/978-3-319-24574-4_28.

Sarver, T., Al-Qaraghuli, A., Kazmerski, L.L., 2013. A comprehensive review of the impact of dust on the use of solar energy: History, investigations, results, literature, and mitigation approaches. Renew. Sustain. Energy Rev. 22, 698–733. http://dx.doi.org/10.1016/j.rser.2012.12.065.

Schepanski, K., Tegen, I., Laurent, B., Heinold, B., Macke, A., 2007. A new Saharan dust source activation frequency map derived from MSG-SEVIRI IR-channels. Geophys. Res. Lett. 34, http://dx.doi.org/10.1029/2007GL030168.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., Ratier, A., 2002. An introduction to Meteosat second generation (MSG). Bull. Am. Meteorol. Soc. 83, 977–992. http://dx.doi.org/10.1175/1520-0477(2002)083<0977:AITMSG>2.3.CO;2.

Schwartz, C.S., Kain, J.S., Weiss, S.J., Xue, M., Bright, D.R., Kong, F., Thomas, K.W., Levit, J.J., Coniglio, M.C., Wandishin, M.S., 2010. Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. Weather. Forecast. http://dx.doi.org/10.1175/2009WAF2222267.1.

Shi, H., Tian, L., Tao, J., Li, Y., Wang, L., Liu, X., 2025. A spatial–temporal deep probabilistic diffusion model for Reliable Hail Nowcasting with radar echo extrapolation. In: Companion Proceedings of the ACM on Web Conference 2025, Association for Computing Machinery. New York, NY, USA, pp. 2367–2376. http://dx.doi.org/10.1145/3701716.3717360.

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

Smith, L.N., Topin, N., 2019. Super-convergence: Very fast training of neural networks using large learning rates. In: Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications. SPIE, pp. 369–386. http://dx.doi.org/10.1117/12.2520589.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S., 2015. Deep Unsupervised Learning using nonequilibrium thermodynamics. In: Proceedings of the 32nd International Conference on Machine Learning. PMLR, pp. 2256–2265. http://dx.doi.org/10.48550/arXiv.1503.03585.

Song, J., Meng, C., Ermon, S., 2021. Denoising diffusion implicit models. In: International Conference on Learning Representations. http://dx.doi.org/10.48550/arXiv.2010.02502.

Srivastava, K., Lau, S.Y., Yeung, H.Y., Cheng, T.L., Bhardwaj, R., Kannan, A.M., Bhowmik, S.K.R., Singh, H., 2012. Use of SWIRLS nowcasting system for quantitative precipitation forecast using Indian DWR data. MAUSAM 63, 1–16. http://dx.doi.org/10.54302/mausam.v63i1.1442.

Trzeciak, T.M., Garcia-Carreras, L., Marsham, J.H., 2017. Cross-Saharan transport of water vapor via recycled cold pool outflows from moist convection. Geophys. Res. Lett. 44, 1554–1563. http://dx.doi.org/10.1002/2016GL072108.

Turner, B.J., Zawadzki, I., Germann, U., 2004. Predictability of Precipitation from Continental Radar Images. Part III: Operational Nowcasting Implementation (MAPLE). J. Appl. Meteorol. 43, 231–248. http://dx.doi.org/10.1175/1520-0450(2004)043<0231:POPFCR>2.0.CO;2.

Wang, Y., Gao, Z., Long, M., Wang, J., Yu, P.S., 2018. PredRNN++: Towards A Resolution of the Deep-in-Time Dilemma in spatiotemporal predictive learning. In: Proceedings of the 35th International Conference on Machine Learning. PMLR, pp. 5123–5132.

Willett, M.R., Brooks, M., Bushell, A., Earnshaw, P., Smith, S., Tomassini, L., Best, M., Boutle, I., Brooke, J., Edwards, J.M., Furtado, K., Hardacre, C., Hartley, A.J., Hewitt, A., Johnson, B., Lock, A., Malcolm, A., Mulcahy, J., Müller, E., Rumbold, H., Rooney, G.G., Sellar, A., Ujiie, M., van Niekerk, A., Whitall, M., 2025. The Met Office Unified Model Global Atmosphere 8.0 and JULES Global Land 9.0 configurations. EGUsphere 1–55. http://dx.doi.org/10.5194/egusphere-2025-1829.

World Meteorological Organization, 2017. Guidelines for Nowcasting Techniques. WMO.

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., Yang, M.H., 2023. Diffusion models: A comprehensive survey of methods and applications. ACM Comput. Surv. 56, 105:1–105:39. http://dx.doi.org/10.1145/3626235.