UNIVERSITY OF LEEDS

Nowcasting Convective Weather: Evaluation, Development and Application of Techniques

Joseph Smith

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Earth and Environment

Institute of Climate and Atmospheric Sciences

June 2025

Intellectual Property

I confirm that the work submitted is my own, except where work that has formed part of jointly authored publications has been included. My contribution and that of the other authors to this work have been explicitly indicated below. I confirm that appropriate credit has been given within the thesis wherever reference has been made to the work of others.

Chapter 4 is based on the following article: J. Smith, C. Birch, J. Marsham, S. Peatman, M. Bollasina, and G. Pankiewicz. *Evaluating pySTEPS optical flow algorithms for convection nowcasting over the Maritime Continent using satellite data.* Natural Hazards and Earth System Sciences, 24(2): 567–582, February 2024. ISSN 1684-9981. doi:10.5194/nhess-24-567-2024. URL https://nhess.copernicus.org/articles/24/567/2024/.

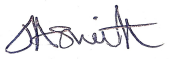
Chapter 5 is based on the following article: J. Smith, C. Birch, J. Marsham, D. Moffat. *SII-NowNet: A machine learning tool for nowcasting convection initiation and intensification in the Tropics.* (Submitted: 2025). Artificial Intelligence for the Earth Systems. American Meteorological Society. (In review)

Chapter 6 is based on the following article: J. Smith, C. Birch, J. Perez, L. Speight, B. Maybee, J. Mooney, J. Marsham. *N-FOREWARNS: Regional nowcasting of surface water flooding with a 3-hour lead time predictability limit.* (Submitted: 2025). Meteorological Applications. Royal Meteorological Society. (In review)

For all of the above articles, the candidate was the lead author, performed the analysis, and wrote the manuscripts. The co-authors were involved in discussing the methodologies and results, suggesting analysis techniques, data collection and making editorial corrections to the manuscripts.

This copy has been supplied on the understanding that it is copyright material and that no

quotation from the thesis may be published without proper acknowledgement.

Signed 

Acknowledgements

First and foremost, I would like to give enormous thanks to Cathryn. Your continuous belief in me and unwavering support have been the key driving forces behind my work and have provided me with opportunities I never imagined possible. Your constant encouragement to travel has given me so many memorable life experiences, for which I will always be grateful. I am incredibly lucky to have had you as my supervisor – you will continue to be my inspiration and friend.

I would also like to say a massive thanks to John Marsham. After stepping in as co-supervisor a year into my PhD, you energised my work at a time when I was feeling a bit lost. Your knowledge of the field and creative ideas have been invaluable in all of our meetings, and you have certainly made me a better scientist. Thanks to Richard Rigby for all your computer wizardry, and David Moffat for your continuous technical support. Your time and input allowed me to achieve so much more than I would have alone. Thanks also to my co-supervisors, George Pankiewicz and Massimo Bollasina, who both gave me great encouragement during my early days.

Thank you to Simon Peatman for providing answers to my relentless onslaught of questions, ranging from cricket white requests to in-depth lessons on MASS commands. You were extremely patient and always spared your own time to help me. Furthermore, to the whole 11.06 office, thank you for all the laughs, tomfoolery, and pints o' cream. It's been an absolute blast. To 'the row' – thanks for the endless inane chatter. Bon bon bon bons...

To everyone involved in the WISER-EWSA project, it was the greatest pleasure to work with each and every one of you. Being part of this project was by far the most meaningful work-related experience I have ever had and has truly inspired my future ambitions. Special thanks to Doug Parker for giving me this opportunity.

Thank you to all of my friends at the Met Office for hosting me during my placement. In

particular, thanks to Katie Norman and the nowcasting team for being so welcoming and involving me in so many interesting discussions. A special thanks to Andrew McNaughton for your hospitality in a time of need! For hosting me during my trip to BMKG, Jakarta – thank you, Achmad Rifani and Agie Putra.

I am so grateful for all the wonderful friends I have made during my time in Leeds. You introduced me to new experiences, and I've learned from you all in so many ways. Special mention to my dearest: Jasper, Eilish, Rachel, Sally, Samir, Ben, Janie and Josefa – thanks for all the endless fun and support. To Louise, always my first port of call and solid rock during the final years: thank you for listening to all of my ramblings, supporting me through the ups and downs, and continuously believing in me. Finally, I would like to thank my parents, Stephen and Vanessa, and my siblings, Issy and Ed, for their continued support and encouragement throughout everything – I am so grateful to have such an amazing family.

Abstract

Convective storms produce hazardous conditions that can lead to natural disasters such as flooding and landslides. Their societal and economic impacts are felt throughout the world, particularly in vulnerable tropical regions. Providing effective early warnings for such events requires accurate short-term prediction — a major challenge in meteorology, especially in the Tropics, where numerical weather prediction models have low skill. Nowcasting fills this capability gap by rapidly generating weather predictions with lead times on the scale of minutes to hours.

This thesis presents advances in convection nowcasting, including satellite-based solutions for the Tropics and the extension of traditional nowcasting techniques to flood prediction. First, traditional nowcasting tools, which are typically radar-based, are applied to the Maritime Continent – a tropical region that experiences regular convective activity – using satellite brightness temperature retrievals, a viable alternative to radar data, which is scarcely available in this region. Overall, these tools demonstrate skill in nowcasting propagating convection up to 4 hours in advance, but struggle to capture the initiation and growth of convection over mountainous regions during the afternoon period. Next, a novel satellite-based machine learning nowcasting tool, SII-NowNet, is introduced. SII-NowNet produces skilful nowcasts of convection initiation up to 2 hours in advance and convection intensification up to 3 hours in advance, over the Maritime Continent. Using Zambia as an example region, SII-NowNet shows that it can generalise well to a previously unseen tropical region without any re-training. Finally, traditional nowcasting techniques are applied to develop N-FOREWARNS, a surface water flood nowcasting tool that generates useful flood risk maps up to 3 hours in advance. N-FOREWARNS is both quantitatively verified and qualitatively assessed by expert users, demonstrating added value to existing operational capabilities.

Overall, the nowcasting developments presented in this thesis show the potential to strengthen early warning systems via improved nowcasting tools and thereby enhance resilience to hazardous weather in vulnerable communities.

Contents

List of Acronyms

ANN	Artificial Neural Network
AUC	Area Under Curve
BT	Brightness Temperature
CAPE	Convectively Available Potential Energy
CIN	Convective Inhibition
CNN	Convolutional Neural Network
DGM	Deep Generative Model
DJF	December, Janurary, February
DL	Deep Learning
EWS	Early Warning System
FAR	False Alarm Ratio
FFC	Flood Forecast Centre
FOREWARNS	Flood fOREcasts for Surface WAter at a RegioNal Scale
FSS	Fractional Skill Score
GPM	Global Precipitation Measurement
IGD	Initiation-Growth-Decay
IR	Infrared Radiation
ITCZ	Inter-Tropical Convergence Zone
JJA	June, July, August
LCL	Lifting Condensation Level
LFSS	Localised Fractional Skill Score
LK	Lucas-Kanade
MAM	March, April, May
MC	Maritime Continent

MCS	Mesoscale Convective System
MJO	Madden-Julian Oscillation
ML	Machine Learning
MOGREPS-UK	Met Office Global and Regional Ensemble Prediction System
MSE	Mean Squared Error
MSG	Meteosat Second Generation
N-FOREWARNS	Nowcasting - Flood fOREcasts for Surface WAter at a RegioNal Scale
NWP	Numberical Weather Prediction
POD	Probability of Detection
POFD	Probability of False Detection
RF	Random Forest
RFG	Rapid Flood Guidance
RMM	Real-time Multivariate Madden-Julian Oscillation
SEVIRI	Spinning Enhanced Visible and InfraRed Imager
SII-NowNet	Simple Initiation and Intensification Nowcasting neural Network
SON	September, October, November
S-PROG	Spectral-Prognosis
SR	Success Ratio
STEPS	Short-Term Ensemble Prediction System
SVM	Support Vector Machine
SWF	Surface Water Flood
WMO	World Meteorological Organisation

List of Figures

List of Tables

Chapter 1

Introduction

In the Earth's atmosphere moist convection begins when, in the presence of water vapour, a parcel of air ascends (freely or mechanically) and cools until it reaches saturation. Further cooling of the saturated parcel triggers condensation and the release of latent heat, which increases the parcel's buoyancy. The continued ascent of the parcel leads to the formation of clouds that can penetrate the full depth of the troposphere.

Moist convection can occur anywhere on Earth; however, it is most commonly observed in the Tropics where high moisture availability, strong surface heating and the large-scale ascent of the Hadley circulation provide favourable conditions for its development. Individual convective updrafts typically occur on scales of <1 km, but when bands of convection develop together, they can form 'organised' convective storms that reach scales >100 km. Due to its chaotic nature, convection is inherently challenging to predict. Small perturbations in initial atmospheric conditions will rapidly amplify over time until a limit of predictability is reached. Furthermore, this limit of predictability is reached quicker for small-scale features (Lorenz 1963; Palmer 2006). Therefore, since convection develops on some of the smallest meteorological scales (with abrupt releases of stored potential energy), predicting it provides one of the greatest challenges in the field.

Organised convective storms generate hazardous weather conditions such as heavy rainfall, high winds, lightning and hail, covering the full spectrum of societal and economic impact levels. Often these storms will initiate and propagate over regions on timescales of hours, leaving little time to react. Low level impact examples include road closures, public transport disruptions

and power outages, whereas high level impacts include flooding and landslides, resulting in destruction of homes, spreading of disease and loss of life. With increased global emissions of greenhouse gases leading to a warming of the Earth, the frequency and severity of impacts from these weather events has and will continue to increase (Calvin et al. 2023). As well as mitigating climate change through reducing these global greenhouse gas emissions, it is also crucial to learn to adapt to living in our changing climate.

The "Early Warnings For All Initiative" was launched by the United Nations in November 2022 and calls for the whole world to be covered by early warning systems (EWSs) by the end of 2027 (World Meteorological Organization 2023). Through this initiative the World Meteorological Organization (WMO) is developing capability for forecasting hazardous weather events, particularly for the world's least-developed countries. EWSs play a key role in being able to adapt to living in a climate that is experiencing an increase in convective storm activity. Although not all impacts are avoidable, EWSs allow time to react and can help reduce the severity of impacts from convective storms. Even a few hours' notice allows time for action to be taken such as evacuating people to safety, closing transport links, preparing back-up power supplies, shutting down airport runways or moving livestock.

A reliable EWS for convective storms requires a skilful forecast that provides enough lead time for decision-making. Convection-permitting numerical weather prediction (NWP) modelling has vastly improved capability in forecasting convection development (Prein et al. 2013; Ferrett et al. 2021). To produce a skilful forecast, these models require atmospheric precursors or existing convective features to be well represented in the initial conditions. However, these initial features are often present on scales smaller than even convection-permitting NWP grid resolutions (Clark et al. 2016), and any small errors in these initial conditions are known to rapidly amplify (Lorenz 1963). Furthermore, the skill at NWP lead times of a few hours (crucial timescales for convective storm development) are problematic due to the required 'spin-up' period (Lean et al. 2008) – the amount of time it takes an NWP model to reach a consistent state after being started from its initial conditions. The skill of NWP is particularly low in the Tropics (Vogel et al. 2020; Zhu et al. 2014), where the weather is dominated by moist convection (Schumacher and Rasmussen 2020) and observational data (required for NWP model initialisation) is sparse. As well as being the wettest region on Earth (Adler et al. 2017), the Tropics is also home to some of the poorest communities in the world, who often don't have the infrastructure to deal with the hazardous

impacts caused by severe rainfall. This vulnerability, combined with low NWP skill, makes the Tropics a key area of focus for developing alternative approaches to convection prediction.

Nowcasting provides a solution to the short-term predictability gap of NWP. The term "nowcasting" was first mentioned in the 1980s with pioneer Keith Browning defining it as a "full description of the weather now, together with extrapolation up to 2 hours ahead" (Browning and Collier 1989). Since then capability has developed to allow skilful predictions at greater lead times than 2 hours, and the WMO updated its definition to "forecasting with local detail, by any method, over a period from the present to 6 hours ahead" (Schmid et al. 2019). However, nowcasting has many different applications (e.g. warning vulnerable communities, aviation, transportation infrastructure and public events), which benefit from predictions with a range of lead times, meaning that limiting the definition of nowcasting to a specific timescale is not always beneficial. Furthermore, as technical capabilities increase, it is likely that lead times will stretch greater than 6 hours. In this thesis, nowcasting is defined as the process of using information on the current atmospheric conditions to rapidly generate short-term predictions of the future atmospheric state.

To fill the skill gap of NWP convection prediction, a nowcasting system must be able to make real-time predictions of small-scale convective processes. This requires analysis of up-to-date, high resolution weather information, which typically comes from observational sources such as weather radars, geostationary satellites or lightning detector networks (or a combination of these). Unlike NWP, nowcasting techniques are not limited by the requirement to solve large sets of complex numerical equations to forecast the atmosphere. Instead, nowcasting requires computationally efficient predictive techniques to produce near-instantaneous, short-term forecasts. To support operational nowcasting, a continuous, efficient workflow of data acquisition is a fundamental requirement of the process.

Weather radar observations are considered the best way to identify the location and intensity of storms over large parts of a country (Saltikoff et al. 2019), which is the necessary input to nowcasting systems. Traditionally, a technique known as optical flow (Horn and Schunck 1981) is applied to radar data to produce nowcasts. Optical flow methods identify clouds and their trajectories through a series of images, and then extrapolate them along the trajectory to generate future predictions (Pulkkinen et al. 2019). These methods show skill at predicting the propagation of convection; however, they are limited to information in the latest radar im-

age, meaning they cannot predict the development of new clouds. Nonetheless, many countries in the mid-latitudes, which are largely covered by a weather radar network, use optical flow for operational nowcasting of convective weather (Bowler et al. 2006; Germann and Zawadzki 2002; Haiden et al. 2011; Marcos 2015). In the Tropics, however, vast regions have unreliable radars or no radar network at all (Permana et al. 2019; Roberts et al. 2022). Geostationary satellite-based observations on the other hand provide a continuous stream of data for monitoring current cloud activity from above (Line et al. 2016). Being able to identify clouds in this way means that satellite-based observations can also be used to produce optical flow-based nowcasts. However, compared to the mid-latitudes, there has been much less focus on the performance of these traditional nowcasting methods (driven with satellite data), presenting a gap in Tropical nowcasting knowledge.

In recent years, Machine Learning (ML) has been the dominant driver of nowcasting developments (McGovern et al. 2023; Prudden et al. 2020). In a process known as 'training', ML models are given large volumes of meteorological data to identify weather features and learn their patterns of behaviour (see section 2.5.2 for full description). Once trained, ML models are applied to unseen data and use their learned knowledge to make future predictions, at near-instantaneous speeds. As ML model architectures have advanced in complexity, they have also been able to account for the chaotic nature of weather by generating probabilistic predictions (Ravuri et al. 2021; Kochkov et al. 2024). This ability to generate rapid, probabilistic output makes ML well-suited to the nowcasting challenge.

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

In the Tropics, ML nowcasting studies have made use of the continuously available satellite data as model input (Han et al. 2019; Lee et al. 2017), producing skill in capturing convection for up to 2 hours on spatial scales of ∼16 km (Lagerquist et al. 2021). However, these approaches still require radar data for model training, resulting in a continued nowcasting gap in the 'radar-less' regions of the Tropics.

Although the development of traditional and AI-based nowcasting tools has improved our short-term prediction of convective storms, predicting their impact is less well understood. Direct nowcasting of rainfall hazards provides the link between crude nowcasting output and understandable, relatable information for end-users (e.g. general public), which is crucial for effective EWSs (Fleming et al. 2015). Many people won't understand the consequences of a high rainfall accumulation until they are directly informed about the hazards that it will bring. Typically, hazard-based nowcasting involves post-processing rainfall amounts to convert them into a warning for a specific type of impact (e.g. flooding). These impact warnings are often discretized into different levels of severity or likelihood of occurrence, making the potential danger clear and easily interpretable.

One of the major hazards caused by convective rainfall is surface water flooding. Also known as a pluvial flood, a surface water flood (SWF) is defined as rainfall-generated water that pools on or flows over the ground surface, before it enters a major watercourse or drainage system. A SWF is often caused by rapidly developing, localised convection, making it an ideal challenge for hazard-based nowcasting (Speight et al. 2021). This need was specifically highlighted in Maybee et al. (2024), who presented a methodology for providing daily regional-scale warnings of surface water flooding. A key outcome of the Maybee et al. (2024) study was that SWF predictions at shorter lead times (0-6 hours), with rainfall nowcasts, could fulfil the end-user need for improved forecast reliability. This need motivates the development of a hazard-based nowcasting tool that can provide rapid, regional-scale information of SWF likelihood.

The overarching aim of this thesis is to improve our understanding and capability for nowcasting convective weather and its hazards. It will firstly address some of the key knowledge gaps in Tropical nowcasting, before focussing on a hazard-based nowcasting approach for surface water flooding.

Chapter 2 summarises the existing literature within the field of nowcasting and Tropical meteorology, whilst chapter 3 will introduce the data and methods that are common to at least

two of the results chapters (data and methods that are specific to only one set of results will be described within the corresponding results chapter). The results chapters will include the following:

i) Chapter 4 evaluates the use of deterministic and probabilistic optical flow nowcasting methods applied to satellite data in the Tropics. Although optical flow is not a novel approach, its performance when applied to satellite data over the Tropics is not well understood. The key research question is:

- Can geostationary satellite data be used for skilful optical flow nowcasting over the Maritime Continent – one of the most convectively active regions in the Topics?

ii) Chapter 5 presents Simple Initiation and Intensification Nowcasting neural Network (SII-NowNet), a newly developed ML-based nowcasting tool that predicts the initiation and intensification of convection in the Tropics. This chapter is motivated by the capability gaps that are highlighted in chapter 4. The key research questions are:

- Can satellite-based ML techniques be used to nowcast isolated convective clouds in the Maritime Continent, and elsewhere in the Tropics?

iii) Chapter 6 presents an application of convection nowcasting for predicting hazardous events. N-FOREWARNS is presented as a tool for nowcasting surface water flooding over England and Wales. Both expert user feedback and quantitative verification is used to evaluate the added value of N-FOREWARNS over current SWF prediction tools. The key research question is:

- Can a SWF nowcasting tool provide added value for England and Wales?

Chapter 7 provides a conclusion of the thesis and discusses the broader implications of the work, its limitations, and future directions.

The beginning of this chapter discussed the correlation between the limit of predictability and spatial scales within convection nowcasting. Figure 1.1 places each of the results sections of this thesis on the nowcasting space-time framework. This is not to act as a performance comparison for each section (they are all addressing inherently different challenges), but to indicate their relative positions in the nowcasting field. The techniques used in Chapter 4 are located towards the top right of the space-time graph, as they are evaluated on the greatest spatial scales and therefore provide the greatest temporal limit of predictability. Chapter 4 will also demonstrate

the challenge of capturing developing convective features, transporting the thesis towards the bottom left of the graph. Chapter 5 will address these smaller-scale challenges, which imposes a lower limit on the temporal predictability. Finally, Chapter 6 moves the thesis to greater spatial scales as it focuses on addressing the challenge of surface water flooding. The small increase in the limit of predictability in this chapter reflects the continuing challenges within this nowcasting domain.

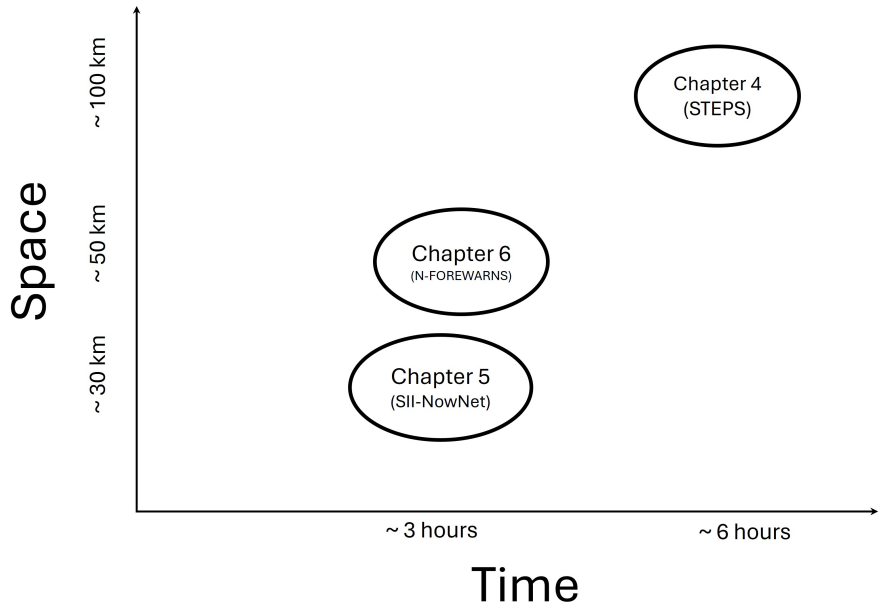


Figure 1.1: A space–time graph of the limits of nowcasting predictability, indicating the location of each chapter (and the model used) within the domain.

Chapter 2

Background and literature review

2.1 Atmospheric moist convection

On a sunny day over land, a parcel of air near the Earth's surface is warmed by direct heat transfer from the ground and moistened through evapotranspiration. This increase in the air parcel's temperature and water vapour content will decrease its density relative to the surrounding air, giving it buoyancy and triggering its vertical motion. The air parcel's journey of ascent depends upon the vertical atmospheric profile, determining whether it will continue rising, become neutral or sink back down. A Tephigram is a thermodynamic diagram used in meteorology to visualise atmospheric profiles (Figure 2.1). Solid grey diagonal lines represent constant potential temperature (from bottom right to top left) and temperature (from bottom left to top right), whilst the dashed near-horizontal grey lines represent different pressure levels. The red line shows the vertical profile conditions of an example environment.

In the example environment, if a parcel of air is lifted adiabatically from the surface (1000 hPa) up to 680 hPa, it will be cooler and denser than the surrounding environment (i.e. negatively buoyant), and thus considered to be stable. Up to 800 hPa the parcel of air will cool at the dry adiabatic lapse rate. At 800 hPa the parcel's temperature drops to a point where the parcel becomes saturated, and water vapour begins to condense – this is known as the lifting condensation level (LCL) (where the temperature equals the dew point temperature). If the saturated parcel is lifted above the LCL, it will condense moisture as it rises and will cool at the moist adiabatic lapse rate (which is dependent on both temperature and pressure). At 680 hPa the environment's atmospheric profile intersects the moist adiabat, known as the level of free

convection (LFC). The LFC signifies the point beyond which the parcel becomes unstable with respect to the environment i.e. it is warmer and less dense than the surrounding air (positively buoyant). The parcel of air will continue to rise to 310 hPa (the equilibrium level), at which point the environmental profile intersects the moist adiabat again and the parcel of air returns to stability.

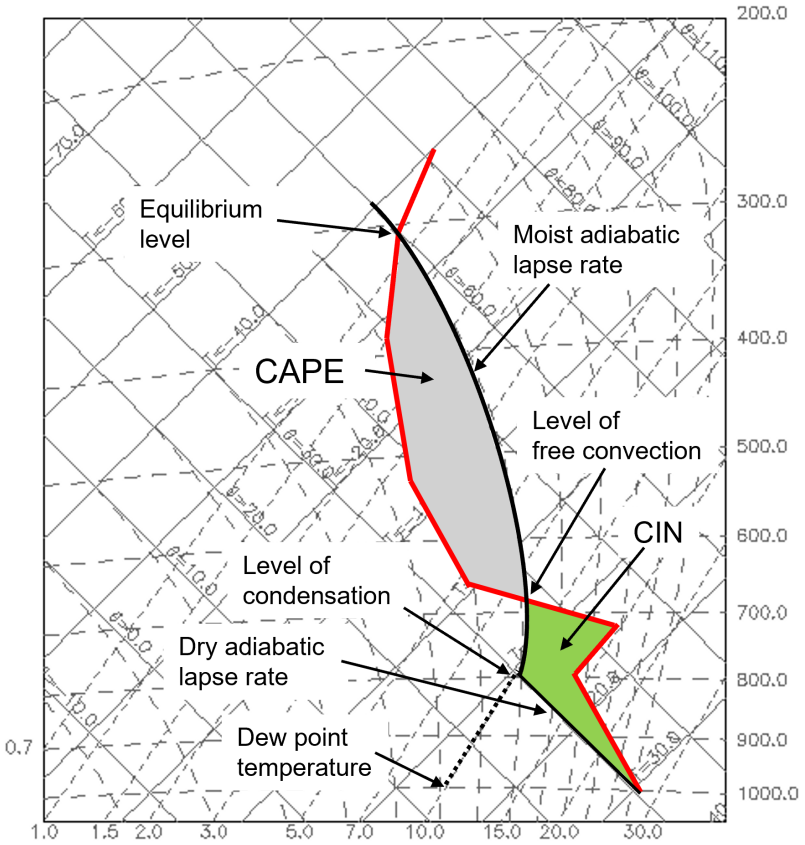


Figure 2.1: A Tephigram of an example environmental profile (red line). The solid black line represents the adiabatic lapse rate that an idealised parcel of air will follow. The shaded regions represent the convective inhibition (green) and convective available potential energy (grey).

In Figure 2.1, the area between the parcel ascent and the environmental temperature profile, to the right of the parcel ascent (shaded green), represents the amount of energy required for a parcel of air to reach the LFC from the surface – known as the convective inhibition (CIN). Above the LFC, the area between the parcel ascent and the environmental temperature profile, to the left of the parcel ascent (grey shaded area), represents the energy available to a parcel of air for ascending (after overcoming the CIN) – this is known as the convective available potential energy (CAPE). Deep moist convection ("cumulonimbus clouds") therefore forms when air can ascend past the LCL and LFC, exploiting the CAPE to reach an equilibrium level near the

tropopause. Higher CAPE results in greater vertical motion (through conversion of potential energy to kinetic energy) i.e. stronger updrafts. This process of deep convection can be further assisted by the mechanical lifting of moist air. For example, the flow of air over orography can cause it to be forced upwards to potentially overcome the atmospheric stability and trigger deep ascent.

When bands of deep moist convection develop in a line or cluster together e.g. squall lines, they can form 'organised' convective storms, which can grow to hundreds of kilometres and last for many days – known as mesoscale convective systems (MCSs). Due their size, MCSs develop their own circulation systems that are greater in scale than any of their component convective clouds. Figure 2.2 provides a diagram of an MCS cross section. Strong updrafts within the convective region of an MCS form towering cumulonimbus clouds, which reach the tropopause before spreading horizontally to form anvil clouds that produce stratiform precipitation (Houze 2004). A key ingredient for the survival of an MCS is vertical wind shear; without it, downdrafts caused by the convective precipitation will collapse onto its updrafts, suppressing the entire system. Wind shear tilts the system enabling the convective downdraft to create a region of cool air on the downshear side of the storm, known as a cold pool. Cold pools are associated with strong winds and a gust front at their leading edge where, crucially, new convective updrafts initiate and merge with the MCS to sustain its lifecycle (Schumacher and Rasmussen 2020).

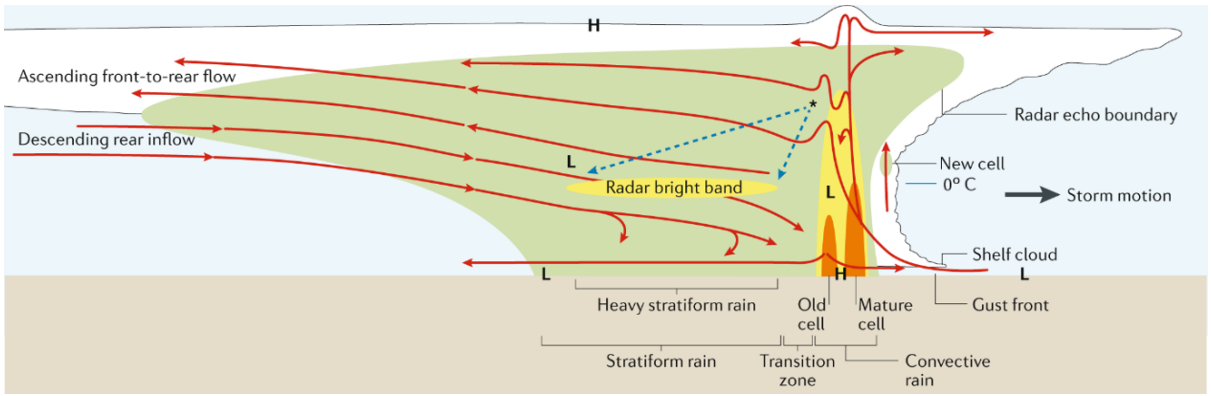


Figure 2.2: A 2-D cross section diagram of an MCS. Taken from Schumacher and Rasmussen (2020).

2.2 Tropical convection

In this thesis the Tropics is defined as the region surrounding the Earth's equator, constrained by the Tropic of Cancer (23.4° N) and the Tropic of Capricorn (23.4° S) (alternative definitions of the Tropics also exist e.g. the region of Earth where total insolation exceeds the outgoing longwave radiation). The intense insolation results in high rates of evapotranspiration and large amounts of moisture present in the atmosphere. Furthermore, the intertropical convergence zone (ITCZ; a band of convergence and large-scale ascent where the trade winds meet) passes northwards and southwards over the equator each year, causing large-scale ascent over the region. The combination of these conditions means that deep convection regularly develops in the Tropics, triggering intense and impactful MCSs. In the Tropics, MCSs contribute to 50-90% of annual precipitation totals (Schumacher and Rasmussen 2020), making them a dominant source of rainfall in the region. Furthermore, MCSs generate a strong upper level circulation that directly contributes to global circulation patterns (Hagos 2010; Schumacher and Houze 2003), meaning they play a crucial role in the Earth's hydrological cycle. Extreme rainfall events from MCSs can have huge societal and economic impacts in the Tropics. Hazards include landslides, flooding, high winds and lightning that often affect less-economically developed areas.

The Tropics is heavily influenced by a seasonal cycle of rainfall, known as monsoons. Monsoons are defined as a change in the prevailing wind direction, which is a fundamental Earth system response to annual variations of insolation (Wang and Ding 2008). In simple terms, Tropical regions generally experience a rainy monsoon when the sun is directly overhead i.e. June, July and August for the northern hemisphere and December, January and February for the southern hemisphere. Greater insolation during these periods drives evaporation rates and land-surface temperature contrasts, enhancing conditions for convection. This simple description provides a basic understanding of the how monsoons are driven; however, in reality the complex nature of the Earth (e.g. land-sea distributions, topography and land types) causes large variations in monsoon characteristics across the Tropics (Chang et al. 2005). This has resulted in the Tropics being split into separate monsoon regions, which are typically: Asia, West Africa, Southern Africa, South America, North America, and Australia (Geen et al. 2020; Zhang and Wang 2008).

The diurnal cycle is one of the dominant components of rainfall variations in the Tropics. A

broad description of the diurnal cycle starts with solar heating of the land during the morning, resulting in peak surface temperatures and near-surface instability around midday. Moist convection begins to initiate in these favourable conditions, which then builds through the afternoon, producing propagating thunderstorms by the afternoon/evening. This results in peak rainfall occurring over the land during the late afternoon/early evening period and over the sea during the early morning (Yang and Slingo 2001). However, complex interactions such as land-sea breezes (J. Slingo et al. 2003) and orographic features (Mori et al. 2004) can cause variations in the typical characteristics of the diurnal cycle. Further investigation into Tropical diurnal cycles has resulted in three defined regimes: oceanic, continental, and coastal (Kikuchi and Wang 2008). The oceanic regime is characterized by peak rainfall during the early morning, whereas the continental regime is characterized by peak rainfall in the afternoon (followed by landward propagation). For the coastal regime, peak rainfall occurs during the afternoon/evening and can propagate both inland and offshore (resulting in two sub-regimes).

2.2.1 Maritime Continent

This thesis has a particular focus on a region of the Tropics known as the Maritime Continent (MC) (Figure 2.3), which contains the countries of Indonesia, Malaysia, Philippines, Papua New Guinea, East Timor, Singapore and Brunei. It is a complex mix of land and ocean with major islands such as Sumatra, Java, Borneo and Sulawesi making it the largest archipelago on Earth. Its complex topography and location across the equator make it a hotspot for storm development. Easterly trade winds blow warm water across the Pacific into the MC creating a 'warm pool' around the region (Dayem et al. 2007), which, when combined with Tropical conditions, further enhances favourable conditions for deep convection – MCSs account for 45 - 70% of total rainfall in the MC (Crook et al. 2024). The large amounts of latent heat released from this convection means that the region is often referred to as the 'boiler box' of the Earth as it plays a crucial role in contributing to the global atmospheric circulations (the Hadley and Walker Cells), in turn affecting both local and global weather systems (Ramage 1968).

Due to the surrounding seas, most islands in the MC experience a diurnal cycle typical of coastal regions (Yang and Slingo 2001). The peak insolation around midday causes a land-sea temperature contrast (due to the lower heat capacity of the land), which drives a sea breeze, triggering convection that builds into the late afternoon and evening. Many islands contain mountainous regions inland, providing an elevated heat source and orographic lift, further enhancing convec-

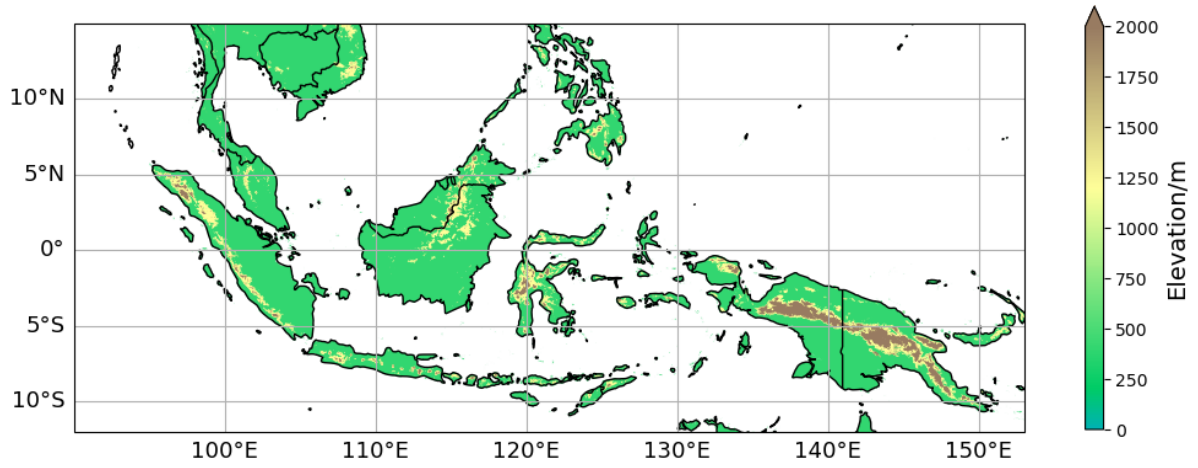


Figure 2.3: Orographic map of the Maritime Continent domain.

tion (Mori et al. 2004). Into the late evening and overnight this convection propagates offshore until the early morning the following day, leaving clear skies over land for strong insolation to restart the process. Figure 2.4 exemplifies the MC diurnal cycle by showing the spatial variation in timings of peak rainfall over the region.

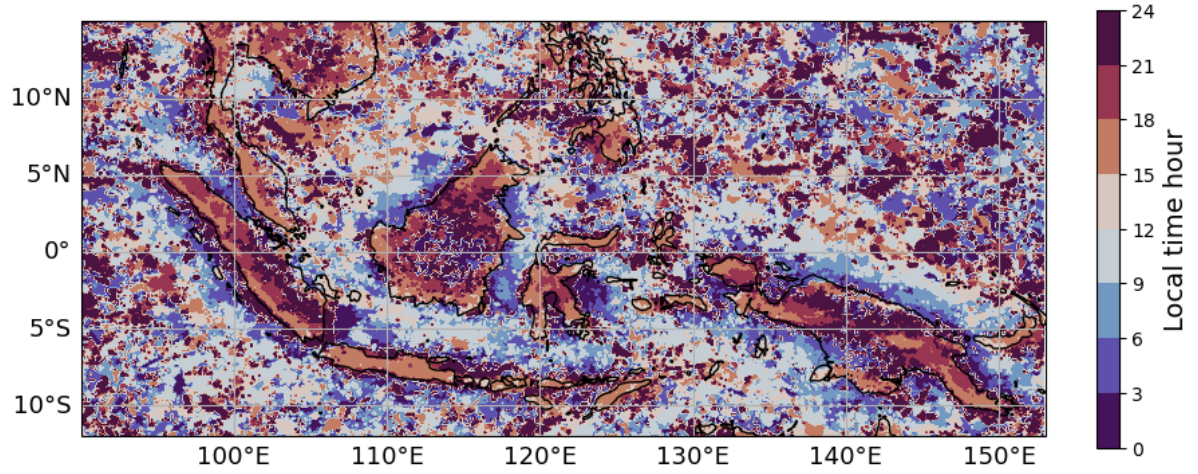


Figure 2.4: The diurnal cycle of peak rainfall over the MC (interpolated to local solar time) using the Global Precipitation Measurement dataset (Hou et al. 2014) from December, January and February 2001 – 2020.

The MC lies at the interconnection between the Asian summer monsoon (June, July, August; JJA) and the Australian summer monsoon (December, January and February; DJF) providing a 'land bridge' for the ITCZ as it swings from the southern to the northern hemisphere during March, April and May (MAM) and vice versa during September, October and November (SON). As the MC islands are spread across both sides of the equator, different regions experience

variations of seasonality (Aldrian and Dwi Susanto 2003; Giannini et al. 2007). Figures 2.5a-d show the mean seasonal rainfall patterns and mean wind vectors as the ITCZ tracks through the MC. Increased rainfall is observed to the north, over East Asia in JJA and to the south, over Australia in DJF. During DJF the majority of the MC land mass experiences high precipitation, whereas in JJA this precipitation is mostly constrained to regions north of the MC (apart from Papua New Guinea which experiences relatively high rainfall all year round). Comparing the distribution of precipitation in MAM and SON highlights the asymmetries between the two transitional seasons. In MAM the band of high precipitation closely tracks along the equator with higher rainfall observed over Sulawesi and Papua New Guinea. In SON, however, the band of high precipitation follows a track further north of the equator with higher rainfall experienced over western Sumatra, northern Borneo and the Philippines. This asymmetry between the transition seasons is likely due to the differences in the location/orientation of the land masses/terrain across the equator, causing asymmetry in land–ocean interactions and hence rainfall patterns (Chang et al. 2005).

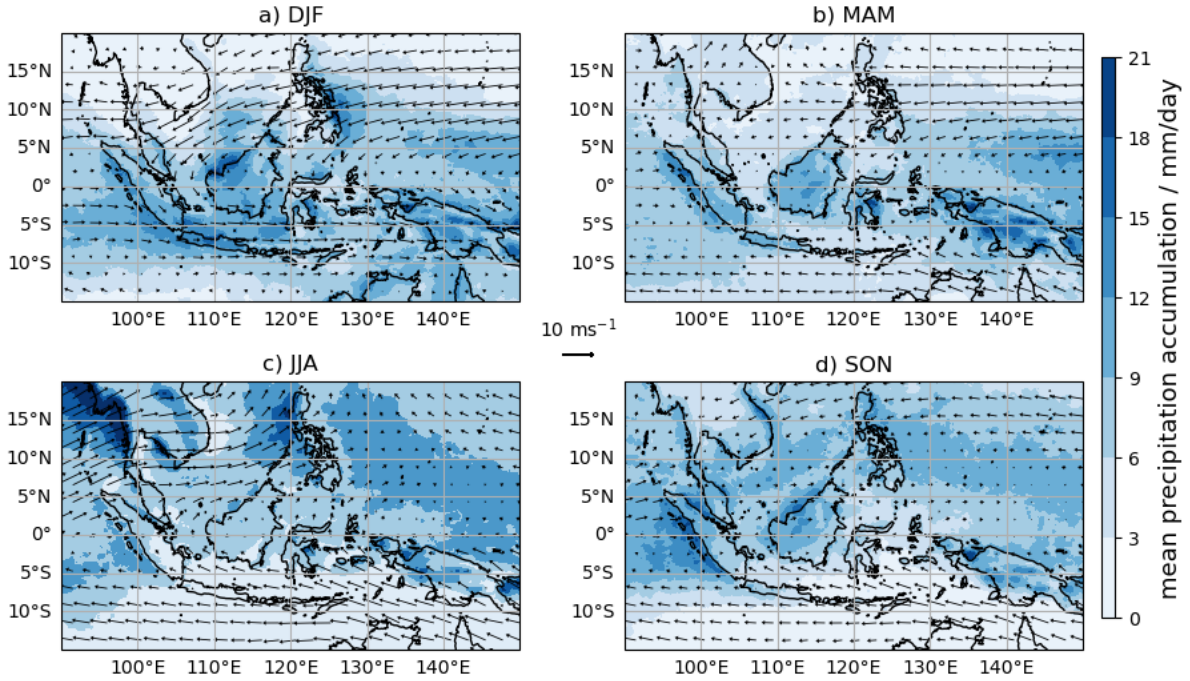


Figure 2.5: Climatological 850 hPa wind vectors plotted with mean daily precipitation accumulation calculated for a) DJF, b) MAM, c) JJA and d) SON. Wind vectors were taken from ERA5 reanalysis (Hersbach et al. 2020) over the period 1979 – 2020. Mean precipitation accumulation was taken from the Global Precipitation Measurement dataset (Hou et al. 2014) measurements over the period 2000 – 2021.

2.2.2 Madden-Julian Oscillation

The leading mode of intraseasonal variability in the tropics is the Madden-Julian Oscillation (MJO), first discovered by Madden and Julian (1971). The MJO is a 30-60 day eastward propagating oscillation in wind, pressure and convection, which moves over the Indian Ocean and into the Pacific Ocean. The MJO is the most dominant source of subseasonal atmospheric variability in the Tropics and is characterised by an alternating active and suppressed phase of strong, deep convection (Zhang (2005)). Active phases of the MJO are associated with higher likelihood of extreme precipitation (Da Silva and Matthews 2021) and an increased amplitude of the convection diurnal cycle over the MC (Peatman et al. 2014).

An MJO event is often divided into eight phases using the real-time multi-variate (RMM) MJO index of Wheeler and Hendon (2004), with each phase indicating the location of the centre of active convection. Figure 2.6 shows how the mean precipitation anomaly patterns are affected during the active and suppressed envelopes of the MJO. As the MJO begins to pass over the MC, it turns from a smooth to a much more complex propagation. The MJO has a strong control on the characteristics of MC convection and where it initiates. For example, a vanguard of precipitation appears over some of the islands prior to the active phase (RMM phase 1) arriving at the western side of the MC e.g. Sumatra. Peatman et al. (2014) hypothesized that this is due to greater insolation prior to cloud coverage arriving from the active envelope, enhancing land-sea contrasts and driving deep convection. Birch et al. (2016) later provided evidence for this by running high resolution regional climate models to analyse the Tropical atmospheric conditions at each phase of the MJO.

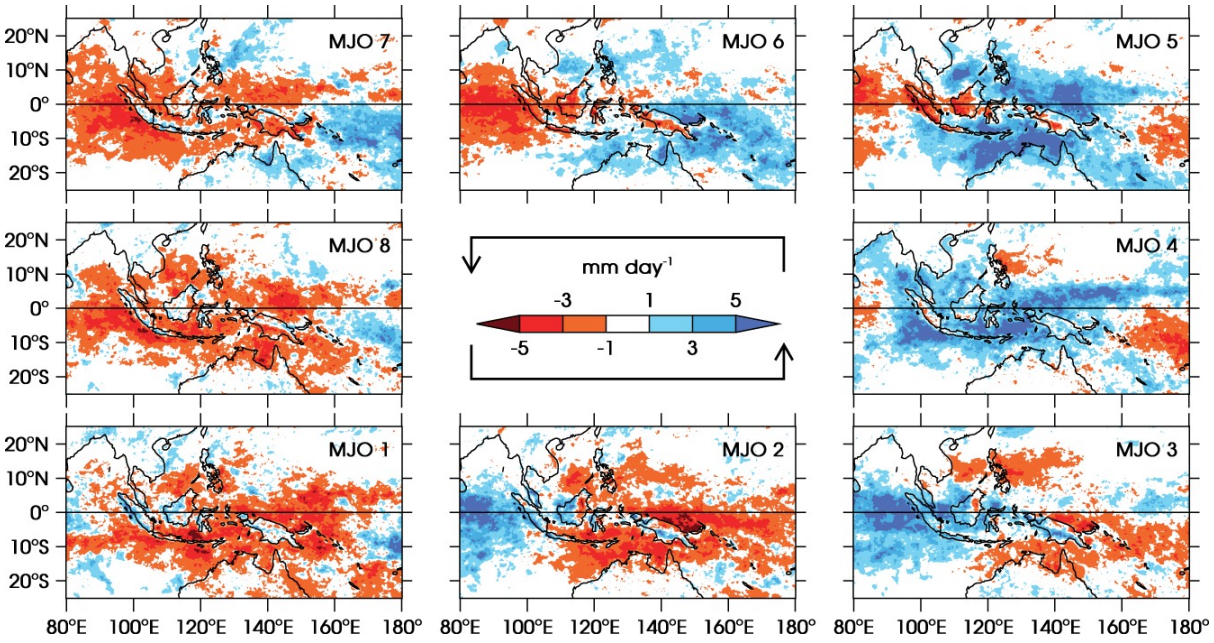


Figure 2.6: Daily mean precipitation anomaly over the MC for each phase of the MJO as defined by Wheeler and Hendon (2004). Phases move forward in time in the anticlockwise direction round the diagram. Taken from Figure 5 of Peatman et al. (2014).

2.2.3 Equatorial waves

Equatorial waves are synoptic scale wave-like disturbances in the atmosphere that travel zonally along the equator. They are 'equatorial' due to the change in sign of the Coriolis force across the equator, which results in trapping them near the equator with their amplitude decaying in the meridional direction (Matsuno 1966). Various different modes of these waves can be observed such as equatorial Kelvin, equatorial Rossby, gravity and mixed Rossby-gravity waves, each with their own pattern of propagation (Masunaga 2009; Wheeler et al. 2000; Yang et al. 2003). Often coupled with convection (N. Takayabu 1994; Straub and Kiladis 2002), equatorial waves play an important role on weather in the MC. High amplitude Kelvin, equatorial Rossby and westward-moving Rossby-gravity waves have been linked to an increase in the likelihood of heavy rainfall events over the MC (Ferrett et al. 2021), whilst their modulation of local wind regimes has been closely tied to a variety of rainfall patterns over the MC islands (Peatman et al. 2021). Active equatorial Kelvin waves have also been shown to modulate the frequency and rainfall totals of MCSs in the MC, with highest rainfall anomalies coinciding with low-level convergence phases (Crook et al. 2024). Overall, through their association with convection, equatorial waves provide a source of predictability for weather in the MC.

2.3 The need for nowcasting

The process of rapidly developing convection can lead to storm formation on timescales of minutes to hours (Peatman et al. 2023). For example, in Indonesia, squall lines are typical convective events that initiate over mountainous regions and propagate across the islands over the course of hours, bringing high winds, rainfall, lightning and destruction along their way (Lo and Orton 2016). This leaves nearby population little time to react and respond to the incoming destructive weather. In order to provide enough warning, there is a crucial need to be able to effectively predict storms on these short timescales and provide rapid, up-to-date information on their development.

NWP models have been successful in revolutionising the way that we analyse and predict the weather. By using approximations to the fluid dynamical laws that govern our atmosphere, NWP models generate descriptions of the future atmospheric state at a range of lead times (Milan et al. 2020). Traditionally, most NWP and climate models rely on a set of parameterisation schemes that simplify unresolved smaller-scale processes (e.g. convection), which have known biases. For example, convection parameterisation schemes are known to produce rainfall that is not organised, too frequent and light and that initiates too early in the day (Berg et al. 2013; Dai et al. 1999). In the MC, coarse-grid resolution NWP models particularly struggle to represent rainfall variations, due to 1) an inability to resolve small-scale features such as mountains (a key influence on the diurnal cycle) and 2) an inability to resolve all islands in the MC and therefore their role in setting up the land-sea breeze (Qian 2008).

More recently the development of much higher resolution models has meant it is possible to explicitly resolve some cloud-scale motions, thus removing the requirement for convection parameterisation (Prein et al. 2013). These convection-permitting models are able to produce more realistic organised convective storms (Prein et al. 2015). Ferrett et al. (2021) evaluated the skill of ensemble forecasts in the Tropics using a convective-scale NWP model at 2.2 km, 4.4 km and 8.8 km resolution. Enabling the model to resolve features at these smaller scales meant that the patterns of convection linked to diurnal variations were better represented. However, the ensemble struggled to produce skill at spatial scales less than $\sim$ 150 km and skill was greatly reduced over night when convection propagates offshore and is no longer constrained by orographic or coastal features. Other convective-scale NWP models applied across the MC have

also struggled in capturing rainfall, often producing it too early and over-predicting accumulation totals (Porson et al. 2019). These limitations are due to initial convective features often being present on scales even smaller than the resolution of the convection-permitting models (Clark et al. 2016). Furthermore, NWP models rely on the assimilation of observations to improve initial conditions estimates and predictions (Hu et al. 2025). The lack of observations in the Tropics (relative to higher latitudes) therefore further compounds these limitations.

After their initialisation, NWP models must undergo a 'spin-up' period. This describes a period of fluctuation before the model reaches a consistent state, and can last multiple hours (Clark et al. 2016; Lean et al. 2008). In Africa, there is evidence that day 2 of NWP forecasts have similar or higher skill than day 1; by day 2 the model has had time to spin-up realistic storms (Woodhams et al. 2018). Organised convective storms can develop in hours, meaning that rapid response is often required within this spin-up timescale, providing a key barrier to the use of NWP models.

The recent increase in skill of machine learning-based weather prediction models (see section 2.5.2 for machine learning details) has led to greater attention being given to these potentially revolutionary forecasting tools. Studies have shown that machine learning-based forecasting can outperform NWP in benchmark statistics such as overall accuracy and root mean squared error (Bi et al. 2023; Kochkov et al. 2024; Lam et al. 2023). Once trained, these models are far less computationally expensive to run compared to NWP methods and don't suffer from spin-up time. However, a comparison study for extreme weather event cases has shown that machine learning-based forecasts do not always outperform NWP (Pasche et al. 2025), highlighting the need for a more comprehensive evaluation of these new tools before fully transitioning them into operationalisation.

In summary, due to their low skill on convective scales and issues with spin-up periods, there is a gap in NWP capability for predicting rapidly developing convective storms. A key solution to this NWP capability gap is nowcasting. Nowcasting is the process of using real-time information on the current atmosphere to generate rapid, short-term predictions of the future atmospheric state. Previous studies have constrained the nowcasting definition to specific lead times, varying between 0 – 6 hours (Browning and Collier 1989; Roberts et al. 2022). However, as the field progresses, methods will produce skilful predictions at greater lead times, requiring a continuously updated definition. Therefore, in this thesis, 'short-term predictions' refers to the

timescales on which convective storms can develop (i.e. minutes to hours) without constraining nowcasting to a specific lead time. Overall, the key essence of nowcasting is the ability to make near-instantaneous predictions of incoming weather to support quick decision-making.

2.4 Nowcasting approaches

For making future predictions of the weather, nowcasting approaches have been broadly split into two methodologies: forward extrapolation and machine learning.

2.4.1 Extrapolation

Extrapolation-based nowcasting methods have been used since the earliest days of nowcasting (Browning 1980). As a broad definition, these methods identify cloud features in images and move them forward in time along motion vectors (extrapolate them) to generate future predictions.

The simplest extrapolation methods assume that a cloud moves along a stationary vector field with a constant level of intensity. Originally, wind fields were used as the motion vectors (Browning and Collier 1989), however, this typically relies on NWP model output, which has known limitations (see section 2.4). As the nowcasting field advanced a technique known as optical flow was adopted from the field of computer vision (Horn and Schunck 1981), and applied to extrapolation nowcasting. Optical flow generates velocity fields to describe the apparent motion of objects across consecutive images. The key assumption of optical flow is that each pixel intensity value remains constant as it is advected along. Given I(x,y,t) is the intensity of a pixel at time t=0, this results in:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \tag{2.1}$$

where Δt is the time between image frames. Applying a Taylor series expansion to equation 2.1 leads to:

$$\frac{\partial I}{\partial x}U + \frac{\partial I}{\partial y}V + \frac{\partial I}{\partial t} = 0 \tag{2.2}$$

where U=$\frac{\mathrm{d}x}{\mathrm{d}t}$ and V=$\frac{\mathrm{d}x}{\mathrm{d}t}$ are the velocity components of the motion field. Equation 2.2 is known

as the optical flow equation. $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$ and $\frac{\partial I}{\partial t}$ can be calculated as they represent the image gradients over space and time, whereas U and V are unknown, meaning equation 2.2 represents an underdetermined system that cannot be solved directly. Optical flow methods attempt to get round this by applying various spatial constraints to U and V (Germann and Zawadzki 2002; Laroche and Zawadzki 1994; Lucas and Kanade 1981) to solve the optical flow equation.

Basic optical flow algorithms provide a fast and efficient methodology for producing nowcasts. However, they are constrained by the unrealistic assumption that the intensity of convection remains the same throughout its lifetime i.e. it only evolves dynamically. Understanding that in reality convection evolves both spatially and dynamically, the Spectral-Prognosis (S-PROG; Seed 2003) algorithm was developed as a step change in extrapolation-based nowcasting. S-PROG accounts for the varying levels of uncertainty at different spatial scales by applying a smoothing function, which increases in intensity as spatial scale decreases. This allows smaller, more unpredictable features to be smoothed out whilst retaining the structure of more predictable, larger features as they are advected along optical flow motion vectors. By highlighting the importance of nowcasting both the spatial and dynamic evolution of convective fields, S-PROG inspired the development of The Short-Term Ensemble Prediction Systems (STEPS; Bowler et al. 2006), a collaboration between the UK Met Office and Bureau of Meteorology, Australia. STEPS is a continuation of S-PROG but with an added function that allows it to capture uncertainties in the advection field at varying lead times too (see section 3.2.2 for full STEPS description). With the injection of varying noise into the optical flow advection field, STEPS is able to generate probabilistic nowcasts for 0 – 6 hours lead times, which aim to account for both the spatial and temporal uncertainty of convection.

STEPS provided a turning point in the progression of extrapolation-based nowcasting. Since then, further tools and methods have been developed for various regions and applications (see Table 2.1). However, there is a key issue that is common with all extrapolation-based approaches: they can only predict the motion of features that are present in the latest image. This means that any initiation, growth or decay of convection will not be captured by these models.

2.4.2 Machine Learning

Today, machine learning (ML) lies at the heart of artificial intelligence. The main basis of ML is setting up computer systems in a way that enables them to learn rules and make predictions without human interference. The overarching concept is simple and highlighted in Figure 2.7. As opposed to the more traditional scientific methods, whereby collected data is inputted into a set of rules (typically in the form of equations), ML systems directly match together data with answers (in a process known as training). During the training process, an ML system 'learns' patterns and behaviours hidden within the datasets. They can then apply these learned rules to unseen data to make predictions (in the same way as traditional methods). As a computer system, it has the benefit of being able to process vast amounts of data very quickly relative to a human. It also has the capability of identifying complex non-linear patterns in data in a way that humans cannot. Overall, the introduction of ML into the scientific community has enabled an alternative approach to solving complex problems (Carleo et al. 2019). In particular, the speed at which ML models can make predictions, makes them a suitable candidate for nowcasting. In recent years, ML has been the dominant driver within the nowcasting field (McGovern et al. 2023).

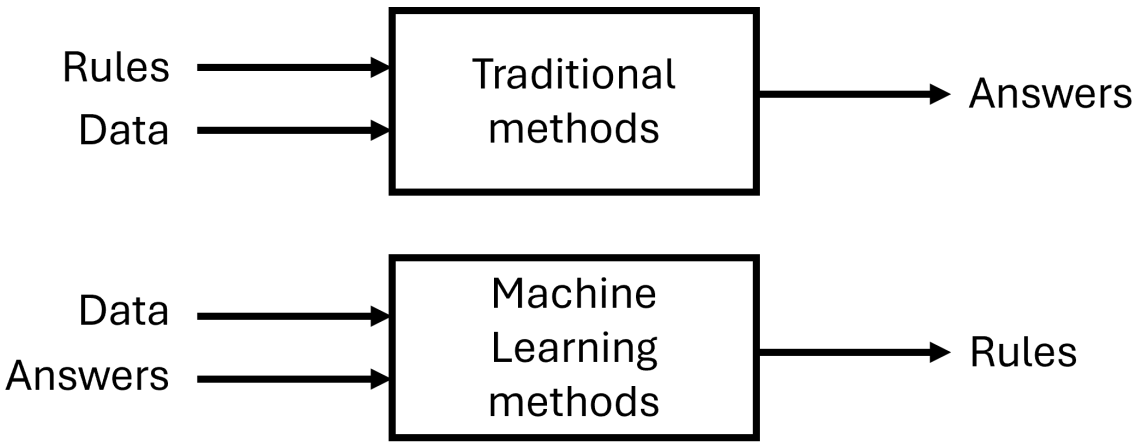


Figure 2.7: A simplistic diagram of the machine learning concept compared to traditional approaches.

Many ML-based nowcasting approaches make use of what are known as 'classical' ML algorithms (Ezugwu et al. 2024). A common choice is the Random Forest (RF) model (Breiman 2001), which is made up of an ensemble of decision trees – tree-like structures made up of decision

nodes – that independently vote on their output. A data point will follow a path through the decision trees (dependent upon its features), until it arrives at a given outcome at the bottom of the tree (the decision) (De Ville 2013). RFs are able to handle high-dimensional data and use their ensemble of decision trees to produce probabilistic nowcast output (Mecikalski et al. 2021). A basic approach was applied by Das et al. (2017) who used RFs to make binary classification predictions of 'rain' or 'no rain' over Calcutta, India, showing RF nowcasting skill for lead times of 30 – 60 minutes. More advanced studies have used RFs to produce spatial maps of rainfall nowcasts by individually predicting the likelihood of convection at each pixel (Lee et al. 2017; Mecikalski et al. 2015). This approach enables local conditions to be considered in the prediction, but can result in impractical computational time when processing individual pixels within high resolution images (Cooner et al. 2016). Furthermore, Mecikalski et al. (2015) showed that a simple logistic regression model (Cole 1991) was able to outperform an RF model, highlighting that there is often more than one ML model that can be applied to a problem.

Another classical ML model that has shown particular nowcasting capability is the Support Vector Machine (SVM; Chih-Chung Chang et al. 2000). An SVM will analyse the features of a dataset to try and derive the optimal way of splitting its data points into distinct classes – particularly applicable to high dimensional datasets. By training on datasets with a range of meteorological features, SVMs have shown skill in predicting radar echoes for lead times of up to 30 minutes (Han et al. 2017; Liu et al. 2022). However, these studies highlighted a tendency for SVMs to produce a high number of false alarms by predicting radar echoes in the incorrect locations.

Further advancements in the field of machine learning have led to the emergence of Deep Learning (DL) architectures. In DL, multi-layered (deep) model networks, known as neural networks, are trained to learn highly complex, non-linear patterns from large volumes of data (LeCun et al. 2015). These types of models are particularly applicable for predicting systems that vary both spatially and temporally (e.g. weather systems), putting them at the forefront of ML weather forecasting research (Kochkov et al. 2024; Lam et al. 2023).

Artificial neural networks (ANNs) are one of the most basic forms of DL architectures (Krogh 2008). Inspired by the structure of the human brain, ANNs are fully interconnected layers of neurons that process information to learn patterns and make predictions. ANNs can be used for both regression and classification tasks and have shown skill in nowcasting the growth and decay

of convection over mountainous regions (Foresti et al. 2019). However, these basic DL models are limited in their capability of dealing with spatial data (Zhan and Datta 2024), which is a crucial requirement when using radar and satellite images. To solve this issue, ANNs have been replaced with convolutional neural networks (CNNs), which use additional filters to convolve features and extract spatial information from images (O'Shea and Nash 2015). A U-Net (Ronneberger et al. 2015) is a type of CNN that is commonly used in nowcasting and has been shown to outperform physics-based precipitation models (Lebedev et al. 2019). Agrawal et al. (2019) first introduced the use of U-Nets for nowcasting classifications of rain types for $0 - 1$ hours lead time. This was developed further by Ayzel et al. (2020), who applied a U-Net approach to nowcast the rainfall intensity values directly. Beyond U-Nets, other nowcasting studies have incorporated a short-term memory element into their CNN-based architectures, which improves capability in remembering important features during iterative predictions of radar frames (Shi et al. 2015), and has shown to better account for the dynamic spatial variability of convection (Shi et al. 2017). Overall, CNN-based approaches have produced skilful nowcasts and outperformed optical flow. However, an issue common within all of these studies is the blurring of predictions as lead time increases (e.g. see Figure 4 of Ayzel et al. 2020).

Increases in compute power have allowed more complex DL models to be trained on high resolution datasets. This has led to the use of deep generative models (DGMs) within nowcasting, which learn probability distributions in the data to generate an ensemble of predictions and capture smaller-scale patterns (Ruthotto and Haber 2021). In order to address the issue of blurry CNN predictions, Ravuri et al. (2021) used a DGM for nowcasting radar images in the UK. They showed that a DGM could outperform optical flow and a standard U-Net model across 80 minutes of lead time. Zhang et al. (2023a) further developed this method by incorporating physical constraints into the DGM nowcasting model to account for the dynamical evolution of precipitation fields. The results of Zhang et al. (2023a) outperformed those of Ravuri et al. (2021), showing greater capability at capturing more intense rainfall regions; however, both DGM approaches still struggled to capture isolated showers and newly developing convection.

More recently the emergence of diffusion models has produced a further step-change in DL capability (Ruiz et al. 2023). These models produce probabilistic output by mapping their inputs to noise fields, and then re-mapping them back again to a target distribution. Although still in their early use for nowcasting, diffusion models have shown improved performance over

DGMs by producing higher quality nowcasts (less blurry) with a more diverse range of ensemble members (Wang et al. 2024; Zhao et al. 2024).

The huge diversity of model architectures, combined with increasing compute power, means that new ML approaches will continue to improve nowcasting skill. However, few studies have focussed on capturing the less frequent (and often more challenging) convective processes, such as initiation, highlighting an open question for further exploration within ML nowcasting.

2.5 Operational nowcasting

Table 2.1 provides information on some of the most advanced operational nowcasting systems globally. These are predominantly based in developed countries, which take advantage of weather radar networks (e.g. STEPS, ANC, S-PROG, MAPLE and SWIRLS) that provide widespread coverage of precipitation. Each system uses a different methodology to suit its application. For example, the INCA system is used for mountainous regions meaning it requires data from a high density weather station network to account for distinct differences in conditions between valleys (Haiden et al. 2011) - highlighting the importance of understanding the local meteorology when developing nowcasting methodologies.

There are 57 radars distributed across Indonesia, Singapore and Malaysia (Permana et al. 2019). However, weather radar networks are expensive to implement, maintain and often not suitable for regions with mountainous terrain (e.g. Sumatra), meaning the types of nowcasting systems listed in Table 2.1 cannot always be implemented in less developed countries. There is, therefore, a widespread interest within nowcasting research in the use of satellite data as the main source of input (e.g. GSWR and NWC SAF – the two most recent systems in Table 2.1), especially in the Tropics. Although often lower resolution than radars, satellites can provide constant, widespread coverage of the Earth's atmosphere from space.

Nowcasting system	**Input sources**	**Region of application**	**Reference**
Short-Term Ensemble Prediction System (STEPS)	Weather radar, NWP	UK	(Bowler et al., 2006)

Integrated Nowcasting System through Comprehensive Analysis (INCA)	Weather radar, NWP, satellite surface station observations	Europe (Alpine regions)	(Haiden et al., 2011)
Short-range Warning of Intense Rainstorms in Localised Systems (SWIRLS)	Weather radar, NWP, satellite, wind profiler, atmospheric sounding, lightning	China	(Li and Lai, 2004)
Auto-Nowcast System (ANC)	Weather radar, NWP	US	(Mueller et al., 2003)
McGill Algorithm for Precipitation nowcasting using Lagrangian Extrapolation (MAPLE)	Weather radar	Canada, US	(Germann and Zawadzki, 2002)
Spectral-Prognosis (S-PROG)	Weather radar	Australia	(Seed, 2003)
Global Synthetic Weather Radar (GSWR)	Satellite, lightning, NWP	US	(Reen et al., 2020)
Support to Nowcasting and Very Short Range Forecasting (NWC SAF)	Satellite, lightning, NWP	Europe	(Marcos, 2015)

Table 2.1: Information on some of the state-of-the-art nowcasting systems that are currently in operational use around the world

2.6 Satellite-based nowcasting

One of the earliest applications of satellite data to weather forecasting came from the TIROS satellite, launched in 1960. Fritz and Wexler (1960) used images taken onboard TIROS's television cameras to produce multiple mosaic images of cloud formations over the northern and southern hemisphere. Although the images have very low spatial resolution, large scale spiral cloud formations can be seen from these early weather satellite images (see Figure 8 in Fritz and Wexler (1960)).

Nowadays, the advancement of weather satellite technology has given us access to data on a much higher spatial and temporal resolution, enhancing our ability to observe small-scale cloud processes (Bessho et al. 2016; Yang et al. 2017). Weather satellites are in a geostationary orbit with an orbital radius of ∼36,000 km and detect a range of spectral radiances emitted by the Earth. By detecting the Earth's radiance at a range of wavelengths, weather satellites can provide information on different atmospheric conditions. For example, over Asia, the Himawari 8/9 geostationary satellite (see section 3.1 for full description) detects wavelengths in the range 0.47 µm–13.3 µm (Bessho et al. 2016). During daylight hours, the visible wavelengths (∼0.47 µm–0.64 µm) allow cloud structures to be clearly observed as they would be with the human eye, and are particularly useful for identifying convergence lines (Birch et al. 2014). The infrared (IR) band (∼3.9 µm–13.3 µm) can be used throughout the day and night, making it highly suited to continuous nowcasting. At IR wavelengths between ∼6.2 µm–7.3 µm, satellites can provide information on water vapour contained within the atmosphere to identify regions that are particularly favourable for storm formation. At greater IR wavelengths (∼10.4 µm–13.3 µm), water and ice (contained within clouds) will absorb the Earth's emitted radiation and re-emit at new radiances. These spectral radiances can be converted to a brightness temperature (BT), which is the temperature that a black body would need to be in order to emit the detected spectral radiance (Petty 2006). BT images of the Earth can be used to easily separate regions of cold clouds from the relatively warmer surface. Rapid scanning modes provide BT images in minutes, allowing forecasters to perform tasks such as assess diurnal land heating trends, accurately track clouds and identify overshooting cloud tops (Line et al. 2016). Overshooting tops are regions of particularly cold cloud temperatures that have penetrated higher than the surrounding cloud. They are associated with strong updrafts and the regions of most intense convection (Bedka and Khlopenkov 2016).

An example of a nowcasting tool that uses multiple satellite channels is Cumulonimbus Tracking and Monitoring (Cb-TRAM; Zinner et al. (2008)). Cb-TRAM is designed for nowcasting convective initiation and monitoring the evolution of convective storms using data from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard the Meteosat Second Generation satellites. Cb-TRAM uses the 10.8 µm, water vapour and high-resolution visible bands to detect and track convective cloud objects from their initial development to the mature phase. Cb-TRAM uses a pyramidal matching technique to derive motion vector fields, which are then used to track cloud objects and extrapolate their movement up to 60 minutes ahead to identify severe weather events. Furthermore, by keeping a log on each convective cell's track history, Cb-TRAM provides a valuable tool for analysing storm characteristics throughout their lifecycles (Merk and Zinner 2013).

Africa provides an example of a region that is benefitting from the implementation of satellite-based nowcasting. Like the MC, it is impacted by intense convective storms but has limited radar network coverage to help identify or track them. Recently, through the Africa SWIFT project (SWIFT 2025), more attention has been focused on developing nowcasting algorithms that make use of geostationary satellite data. Testbed events during the SWIFT project have provided the opportunity to test these algorithms, whilst working with local experts to improve impact-based nowcasting in operation (Fletcher et al. 2023). One of the key outputs of the SWIFT project is the Forecasting African Storms Application (FASTA), which is a publicly available application that provides live and up-to-date information on storm activity over Africa (using satellite data).

2.7 Tropical nowcasting

The current distribution of geostationary satellite orbits around the Earth means that the full Tropical belt is within line of site of at least one satellite, meaning weather data is available across the Tropics (Bessho et al. 2016; Schmetz et al. 2002; Schmit et al. 2008). With vast areas of the Tropics having unreliable weather radar networks (Permana et al. 2019) or no network at all (Roberts et al. 2022), geostationary weather satellites are a key source of input data for nowcasting tools (particularly operational systems).

The Meteosat satellite series provide a key source of data for African nowcasting (Schmetz et al. 2002). The NWCSAF consortium (Table 2.1) sits within the EUMETSAT satellite applications

facility (Holmlund et al. 2025), providing nowcasting products driven by Meteosat satellite data. One of these products is the Convective Rainfall Rate (CRR; Marcos 2015), which uses IR, water vapour and visible channels to predict current rainfall rates – known as rainfall retrievals – across Africa. Extrapolation schemes (using cloud feature displacements) are also applied to produce nowcasts up to 90 minutes ahead. Over the Guinea coast, CRR showed useful skill for the full 90 minute extrapolation, with greater skill being shown during the day when visible channels are also available (Hill et al. 2020). Another African NWCSAF product is the Rapidly Developing Thunderstorm (RDT; Autones 2016) algorithm. The RDT uses IR channels to track convective cells through time and associate them with a "phase of development" and "severity" attribute. An extrapolation scheme is used to produce predicted cell tracks for up to 60 minutes ahead. Over the Guinea coast, the RDT detected 60% of rainfall events with rain rates of >5 mm/hr, with this percentage increasing with greater rainfall intensity thresholds (Hill et al. 2020). Further testing over South Africa, showed that improved RDT performance could be achieved when incorporating lightning data as an auxiliary input (Gijben and de Coning 2017).

Aside from the NWSCAF consortium, a range of other satellite-based nowcasting techniques have been tested over Africa. Burton et al. 2022 applied the Lucas-Kanade optical flow algorithm (Lucas and Kanade 1981; see section 3.2 for full description) to satellite rainfall retrievals to produce nowcasts for lead times of up to 6 hours. They showed it had skill on a 10 km scale for 2-hour lead times; however, as with all optical flow methods, it was unable to capture newly initiating and developing convection. More recently, an ML-based rainfall retrieval tool has been developed for Africa, known as Rain Over Africa (RoA; Amell et al. 2025), which predicts rainfall rates when inputted with BT data. Qualitative comparison testing in a recent African nowcasting testbed (Ageet et al. 2024) suggested that RoA could produce more accurate rainfall information than CRR retrievals. Currently, RoA is limited to retrievals only, but further work is currently underway to incorporate future rainfall predictions too (Amell, personal communication, 2024). In Africa, an alternative source of predictability is soil moisture (Taylor et al. 2012). Taylor et al. (2022) found that, by using land surface temperatures in satellite observations as a proxy to soil moisture, positive land surface anomalies have a strong link to intense convection in storms over the Sahel. This key result highlights the importance of surface temperature observations and the potential added skill they can provide for future nowcasting tools in Africa.

The few nowcasting methods that have been tested for Southern and Central America have taken advantage of data from the Geostationary Operational Environmental Satellite (GOES; Schmit et al. 2008). Islas-Flores and Magaldi (2024) applied the Lucas-Kanade and S-PROG methods to a region of Mexico, and showed that both had a ∼70% likelihood of predicting heavy rainfall events with a 15 minute lead time. Further south, Ribeiro et al. (2019) presented a case study of a severe thunderstorm over Southeastern Brazil to evaluate the added-value of the GOES rapid scan (full domain scan at intervals down to 1 minute) and highlighted the advantage of increased temporal resolution for extending nowcast lead times. However, large oscillations in the variables of the 1 minute interval scan resulted in a recommendation of the 5 minute scan interval for nowcasting. Another case study evaluation was performed by Seal et al. (2024), covering a storm that propagated over Colombia, Ecuador and Peru. They used a DL model to predict the next frame of satellite data for 4 hours of lead time, showing skill for the first 2 hours. Further testing of this model over a larger number of samples is required for a more robust result, but it provides a good baseline for further ML-based nowcasting developments over a region that has received relatively little attention in this field.

Tropical Asia is home to a large proportion of the world's population and receives the majority of its rainfall from convective storms (Schumacher and Rasmussen 2020), making it a key focus for nowcasting studies. Unlike most Tropical countries, India has a significant number of weather radars available to use for convection nowcasting (Sen Roy et al. 2014). SWIRLS (Table 2.1) and the Warning Decision Support System (an extrapolation-based tool) were tested on radar data over the cities of Delhi and Calcutta, with both studies highlighting a need to develop tools that better capture the growth and decay of storms (Lakshmanan et al. 2007; Srivastava et al. 2021). Goyal et al. (2017), on the other hand, applied a cloud tracking algorithm (Vila et al. 2008) to BT data to extrapolate intense storms over India, producing reasonable skill for a 30 minute lead time. Other satellite-based nowcasting approaches in Asia have made use of the Himawari 8/9 geostationary satellite. In Korea, Lee et al. (2017) trained an RF model to identify convection initiations (identified in radar data) using Himawari 8/9 BT data and captured ∼80% of the observed events (at the expense of a high number of falsely predicted events). Han et al. (2019) took this approach further by tracking convective cells back to their origin, to generate a more rigorous dataset of convection initiation events. They also trained an RF model on this new dataset, which produced a greater proportion of falsely predicted events than Lee et al. (2017), but overall provided a better reflection of the challenge in predicting isolated initiation events.

More advanced approaches have applied DL models for predicting convective events in Tropical Asia. Typically, these models identify convective features in satellite imagery by training them on convective features observed in radar images (Lagerquist et al. 2021; Sun et al. 2023). Once trained, solely the satellite data, which has far greater coverage than radar, can be used for predicting convection. Overall, these DL approaches have shown the greatest advancements for nowcasting Tropical convection in Asia, but they often lack skill beyond a 2 hour lead time and cannot be trained over the 'radar-less' regions.

Over Southeast Asia, Naha Biswas et al. (2024) utilised atmospheric gradients (satellite signal delays caused by atmospheric moisture content) to indicate incoming precipitation with a ∼12 hour lead time. These gradients were used as input to a U-Net model, which demonstrated skill capturing rainfall within a 6 hour period on a 8°x8° spatial scale. Traditional optical flow methods have also been applied over the region by Thu et al. (2025). They utilised the Vietnam's radar network for 1 - 6 hour lead time nowcasts, highlighting particular challenges of capturing convection over the complex orography. For the development of an early flood warning system in Vietnam, La et al. (2024) used Himawari BT data and surface wind speeds over to characterise the severity of convective systems. However, the use of low earth orbiting satellites to derive wind speeds means that this system is not yet operationally viable. The MC itself has received little attention in the field of nowcasting, despite the region experiencing regular intense convective activity, affecting the lives of millions. The Indonesian network consists of 42 weather radars (Permana et al. 2019) but is sparse relative to the size of the country; the country is highly mountainous and experiences communication issues between sites, meaning real-time full radar coverage of the region is not possible (Permana et al. 2019). One of the radars within the network was used by Ali et al. (2021) to nowcast two rainfall events over southern Borneo. On the other hand, satellite data were used by Harjupa et al. (2022) to apply the Rapidly Developing Cumulus Area algorithm (Sobajima 2012) to a region of western Java for predicting 77 heavy rainfall events. The limited sample size and domain of these studies makes it difficult to understand how effective the methods are for other regions of the MC.

Nowcasting is an active field of research across the Tropics with a variety of tools and approaches being tested. Although some studies have specifically focussed on nowcasting newly initiating convection, their continued use of radar data creates a barrier to application in other regions. There is, therefore, a need for a purely satellite-based nowcasting system that enables widespread

application.

2.8 UK convection

Due to its location at higher latitudes, the UK receives far less rainfall than most parts of the Tropics. Weather in the UK is largely influenced by frontal systems, which bring widespread cloud cover and relatively light, stratiform rainfall. In addition to this, synoptically driven convection can occur when larger-scale weather systems, such as mid-latitude cyclones, create conditions that encourage localized upward motion of warm air. This can lead to the development of cumulus clouds and heavy showers or thunderstorms (Lewis and Gray 2010). While less frequent than rainfall associated with frontal systems, these convective events contribute to the variability of UK rainfall and can produce more intense precipitation over localized area (Rüdisühli et al. 2020).

A 20-year daily rainfall average over the UK shows peak rainfall during the winter months (December, January and February) and minimum between March – July (Figure 2.8). Although summer months experience less rainfall on average compared to winter, they are often more susceptible to extreme rainfall events (Hand et al. 2004). During summer, an increase in temperature results in favourable conditions for convective storm development, which produces large downpours and higher than average rainfall amounts. This convectively driven extreme rainfall plays a dominant contributor to disastrous hazards in the UK such as flash floods and destructive winds (Davies 2017). Furthermore, a study by Jones et al. (2013) showed that extreme rainfall events in the UK increased in both their intensity and duration for the period 1961 – 2009. Although there is good understanding of the theory behind UK convective storm development (Bennett et al. 2006), realistically representing this in forecasting models remains a challenge (especially on the smallest scales), providing a major focus for UK meteorological research (Flack et al. 2023) and field campaigns (Barrett et al. 2021).

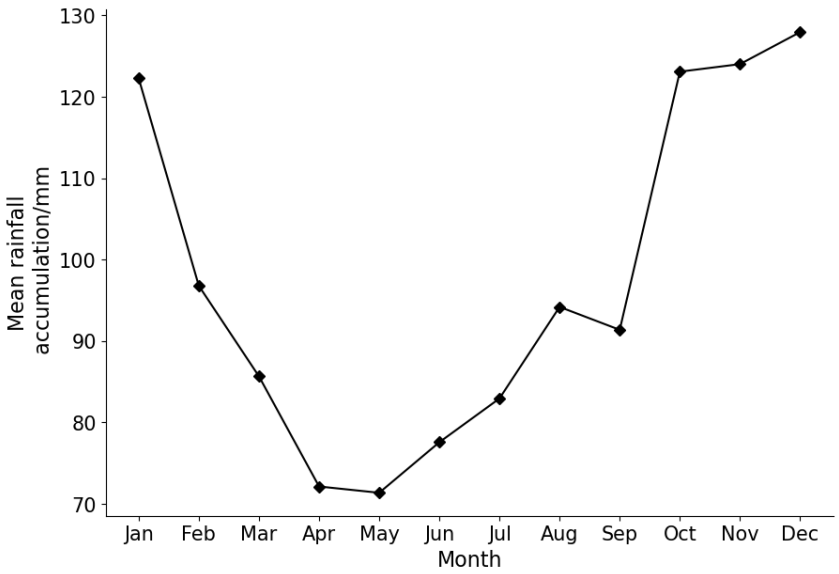


Figure 2.8: Average UK monthly precipitation amounts from the period 1991 – 2020. Data taken from the Met Office HadUK-Grid dataset of gridded climate observations (Hollis et al. 2019).

2.8.1 UK nowcasting

Currently, the UK Met Office uses a STEPS-based system (see section 3.2.2 for further description) for operational short-range forecasts. The STEPS system is fully automated and blends extrapolated radar rainfall nowcasts (0 – 2 hours) with NWP data (up to 24 hours) to generate rainfall predictions across the UK (Seed et al. 2013). Rainfall data from the blended STEPS system is also used to drive an operational hydrological forecast models (Price et al. 2012). However, nowcasting is an area of continued research at the UK Met Office, with new techniques currently under development such as the PLUVIA Cell Tracker and the PLUVIA Mesoanalysis tool (Nowakowska and Smith 2024).

2.8.2 UK surface water flooding

The nowcasting methods that have been discussed so far in this chapter have made direct predictions of meteorological variables such as rainfall rates, radar reflectivity or BTs. However, these predictions do not provide information on the hazards that occur due to rainfall. This is a crucial, yet often overlooked, part of end-to-end nowcasting, providing users with interpretable and actionable information.

Surface water flooding, or pluvial flooding (Falconer et al. 2009; Speight et al. 2021), is typically caused by intense convective weather events that produce high volumes of rainfall over localised

regions, in short periods of time. Due to the rapid onset of these events there is often little time to prepare for surface water flooding, which results in severe impacts to local communities (Carlyon et al. 2012). It is estimated that 4.6 million properties in England are at risk of surface water flooding (Environment Agency 2025). The Coverack, Cornwall flood on 18th July 2017 provides an example of the impacts that surface water flooding can cause, during which an estimated 165–201 mm of rainfall fell in 3 hours (Flack et al. 2019), destroying many houses in the village with flowing water up to 1 metre in depth (Baker 2017). The event was caused by an extremely localised convective storm over Coverack – the nearby weather gauges didn't record any significant rainfall. Improved early warning that the event might happen, even an hour ahead of the event, would have at least allowed the emergency services to be on standby for a quicker response. Nowcasting methods provide an opportunity for improving SWF prediction for rapid emergency response.

2.8.3 SWF prediction approaches

A wide range of methodologies have been developed for SWF forecasting, for urban to national scales (Speight et al. 2021). The least complex methods use empirically-based rainfall thresholds to derive SWF warnings (Acosta-Coll et al. 2018; Javelle et al. 2016). Warnings are triggered if rainfall observations (e.g. rain gauges or radar estimates), exceed pre-defined depth-duration thresholds. In the US, the Flash Flood Guidance product is used to determine the amount of runoff generated from rainfall totals. Different rainfall thresholds are used across the country depending on the local geography e.g. mountains or flood plains, to generate accurate indications of local flash flood occurrence (Clark et al. 2014). In the UK, the Met Office developed the Extreme Rainfall Alert (ERA) system, which provides warnings when extreme rainfall events are likely to cause severe SWFs in urban areas (Hurford et al. 2012). The ERA system combines city-specific rainfall thresholds with storm durations to identify appropriate warning levels for SWF alerts. Although the use of empirically based rainfall thresholds offers a good initial warning estimate and low run costs, their simplicity limits the spatial detail of warnings, and they cannot account for the dynamically changing conditions that can affect SWFs e.g. ground saturation levels.

To incorporate more detail into SWF warnings, output from real-time forecasting models is combined with predefined flood impact scenarios, linking rainfall to associated flood impacts. These approaches can account for the spatial variability of rainfall fields, whilst still maintain-

ing low computational expense (Speight et al. 2018). In France, the Adaption of Geographic Information for Flood Warning system uses a hydrological model with a 1 km grid spacing that converts real-time radar rainfall into magnitudes of water flow (Saint-Martin et al. 2016). By overlaying land use maps, it produces an overall flooding exposure index that identifies regions most at risk from surface water flow. Dottori et al. (2017) applied a similar method across Europe by producing flood risk forecasts with a continental-scale hydrological model. Combining this with vulnerability and exposure maps of major cities across Europe, they mapped flood volumes to classify the regions that are most at risk of impacts. In the UK, the Surface Water Flooding Hazard Impact Model (SWFHIM) is currently in operational use at the Met Office (Aldridge et al. 2020). SWFHIM uses hazard, impact and risk mapping models to convert ensemble rainfall forecasts into likelihoods of minimal, minor, significant and severe SWF impact levels (see section 6.2.3 for full description of SWFHIM). More recently, Flood fOREcasts for Surface WAter at a RegioNal Scale (FOREWARNS) has been developed for England and Wales (Maybee et al. 2024), which generates flood risk maps by comparing reasonable worst case rainfall scenarios with pre-defined look-up thresholds from hydrological modelling (see section 6.2.3 for full description of FOREWARNS). FOREWARNS is currently being tested for operational use by the UK Met Office to work alongside SWFHIM.

Although the use of predefined flood impact scenarios has enabled more detailed flood warnings, urban-scale flood modelling often requires even finer resolution. To overcome this, real-time hydrodynamic simulations are coupled with hydrological models to produce high resolution warnings, on street-level scales (Green et al. 2017). Guerreiro et al. (2017) demonstrated the use of urban hydrodynamic modelling across 571 European cities for different rainfall events. They were able to provide flood depths at a 25 m resolution and demonstrated that, on these scales, the models are highly sensitive to urban elevation, topography and water flow paths. The requirement for high spatiotemporal resolution rainfall data also opens the opportunity for SWF nowcasting (Willems 2016). In Leicestershire, UK, Yu and Coulthard (2015) applied a simple urban hydrodynamic model to a case study flood event, proving its predictive capability on hourly lead times. However, high computational costs and the long run-times required for these approaches mean that these models are not yet suitable for operational SWF forecasting/nowcasting for national-scale warnings. Nevertheless, as computational capacity increases it will be possible to run hydrodynamic simulations on greater spatial scales, improving our capability for surface water flood prediction.

The broad range of SWF prediction methodologies discussed in this section highlights the continued focus on development within the field as well as the need to effectively capture SWF events. Indeed, a key outcome of the Maybee et al. (2024) study was that, given the challenges of forecasting rapidly intensifying convection with NWP, producing SWF predictions at shorter lead times (0-6 hours) using rainfall nowcasts, would fulfil the user-need for reliable SWF forecasts. However, thus far, there has been no development of a SWF tool that focuses on these key timescales. There is therefore a need for a nowcasting SWF prediction tool that has operational capability in the UK.

Chapter 3

Data and methods

The data and methods described in this chapter are common to at least two of the results chapters (as referred to within the text). Any data or methods that are unique to only one of the results chapters will be described within the results chapter itself.

3.1 Geostationary satellite data

Chapters 4 and 5 use BT data from geostationary satellites as input to nowcasting algorithms (see section 2.7 for full details on geostationary satellites). Convective clouds can be identified in geostationary satellite BT images from their colder tops relative to the surface of the Earth.

The Himawari-8/9 geostationary satellites are located at 140.7°E above Earth. Himawari-8 was the primary operational satellite between 2015 – 2022 (with Himawari-9 in orbital-standby), before Himawari-9 took over on 13th December 2022 (Zhu et al. 2024). The Advanced Himawari Imager (AHI) onboard Himawari-8/9 has 16 band channels ranging from 0.47 to 13.3 µm, covering parts of the visible, near-infrared and infrared (IR) spectrum (Bessho et al. 2016). BT data from channel 13 on board the Himawari-8/9 AHI has been used in this thesis, which detects IR radiation with a wavelength of 10.4 µm. The 10.4 µm BT images are transformed to a Cartesian grid with a grid spacing of 2 km, using the gdalwarp command from the Geospatial Data Abstraction Library (Rouault et al. 2023).

The Meteosat Second Generation (MSG) satellite series are a set of European geostationary satellites. The Spinning Enhanced Visible and InfraRed Imager (SEVIRI) on board each satellite observes the Earth in 12 spectral channels ranging from 0.64 µm to 13.4 µm. Chapter 5 uses

the Meteosat-10 satellite, which is located at 0° above Earth and detects IR radiation at a wavelength of 10.8 µm, every 15 minutes (Schmetz et al. 2002). The 10.8 µm BT images are transformed to a Cartesian grid spacing of 3 km.

3.2 Optical Flow algorithms

Chapters 4 and 6 of this thesis use nowcasting data generated by the Lucas-Kanade algorithm (Lucas and Kanade 1981) and the Short-Term Ensemble Prediction System (Bowler et al. 2006).

3.2.1 Lucas-Kanade

In this thesis BT fields are inputted to the Lucas-Kanade (LK) algorithm to produce nowcasts. The LK algorithm is an optical flow method that assumes, for a given pixel, the eight immediately surrounding pixels move along with that given pixel. This assumption results in nine separate versions of the optical flow equation (equation 2.2) (eight from the surrounding pixels and one from the given pixel itself), representing an overdetermined system. A least squares fit method is then applied to the nine equations to obtain the optimum solution for the given pixel. To create a motion field, the LK algorithm first identifies key features within three images spaced in time (e.g. T – 2 hrs, T – 1 hrs and T – 0 hrs, where T is the current time) by using the Shi–Tomasi corner detection algorithm (Jianbo Shi and Tomasi 1994). The velocity field components are then calculated for each feature (using the LK assumption) to create a sparse motion field (velocity vectors for features only). The feature velocity vectors are then interpolated across the whole field (where there are missing vectors) to generate a dense motion field (Figure 3.1). Once the motion field has been determined, the latest observation (T – 0) needs to be advected along the motion field. The nowcast is the resulting advected field.

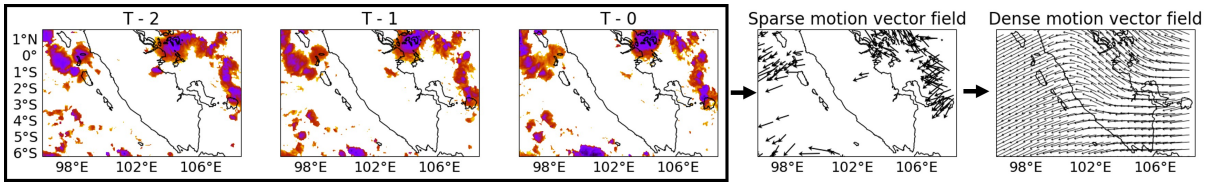


Figure 3.1: An example (using BT data over Sumatra) of how a motion field is generated using the LK method. Features are identified and tracked across the three input images (T-0, T-1 and T-2) to generate the sparse motion field. The sparse motion field is then interpolated onto the rest of the domain to produce the dense motion field.

3.2.2 Short-Term Ensemble Prediction System

The Short-Term Ensemble Prediction System (STEPS) was jointly developed by the UK Met Office and the Bureau of Meteorology Research Centre, Australia, and aims to address the issue of unpredictability in the lifetime of convection by injecting fields of varying stochastic noise. Firstly, STEPS is inputted with observational fields such as radar rainfall or BT. It applies a fast Fourier transform to the latest meteorological observation field to decompose it into cascades of different length scales. Varying intensities of Gaussian noise fields are then injected into each cascade field depending on the length scale. Cascades containing the small length scale features will receive greater intensity of noise injection, as these features represent the greatest uncertainty in growth and decay of convection. In contrast, the large length scale features receive a lower intensity of noise, as these features represent the least uncertainty in growth and decay. The cascades are then recomposed to produce a modified observation field, which is ready for extrapolation using the LK optical flow algorithm.

To capture the uncertainty in the extrapolation, STEPS also applies noise perturbations to the motion vector fields. The magnitude of the perturbation increases with respect to lead time as the motion field increases in uncertainty. The modified observation field is extrapolated along the perturbed motion field to create one ensemble member of the nowcast. Ensemble members are generated by using new realisations of the noise perturbations to create multiple versions of the nowcast. Chapter 4 produces 20-member ensemble STEPS nowcasts by using solely geostationary satellite BT data.

STEPS is also able to blend its extrapolated fields with other sources of forecast data. In Chapter 6, radar rainfall fields (see section 6.2.2 for radar rainfall description) are initially extrapolated using the STEPS algorithm for up to $\sim$ 3 hours' lead time. Beyond this lead time the algorithm transitions to using downscaled rainfall output from the 12-member UK regional Met Office Global and Regional Ensemble Prediction System (MOGREPS-UK, Porson et al. 2020). MOGREPS-UK is a convection-permitting ensemble NWP model that produces forecasts for the UK and Ireland every 5 minutes with a 2.2 km resolution. Overall, after the blending and noise injection stages, STEPS produces a 24-ensemble member nowcast (12 ensemble members from MOGREPS-UK plus 12 ensemble members from noise injections) that covers the UK with a resolution of 2.2 km, every 15 minutes, for lead times up to 6 hours.

3.3 Verification

3.3.1 Persistence nowcast

The nowcasting tools presented in Chapters 4 and 5 are evaluated against persistence nowcasts. A persistence nowcast is a simple prediction method, which assumes that the current weather will persist into the future i.e. the weather now will be identical at the next prediction timestep. This baseline model is produced by using the latest observation as the next prediction for a given lead time.

3.3.2 Basic verification statistics

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

These metrics are used to calculate skill scores in all three results chapters. The accuracy is given by:

$$accuracy = \frac{a+d}{a+b+c+d} \tag{3.1}$$

which captures the overall fraction of correctly predicted events/no events.

The probability of detection (POD) is given by:

$$POD = \frac{a}{a+b} \tag{3.2}$$

which captures the proportion of observed events that were correctly nowcasted. A POD value of 1 represents a perfect score and 0 represents the least skill (in terms of POD).

The false alarm ratio (FAR) is given by:

$$FAR = \frac{c}{a+c} \tag{3.3}$$

which captures the proportion of nowcasted events that were false alarms. A FAR value of 0 represents a perfect score and 1 represents the least skill (in terms of FAR).

The probability of false detection (POFD) is given by:

$$POFD = \frac{c}{c+d} \tag{3.4}$$

which captures the proportion of observed 'no events' that were nowcasted as events. A POFD value of 0 represents a perfect score and 1 represents the least skill (in terms of POFD).

The success ratio (SR) is given by:

$$SR = \frac{a}{a+c} \tag{3.5}$$

which provides the likelihood of an observed event, given that an event is nowcasted. This skill metric provides information on the balance between false alarms and hits. An SR value of 1 gives a perfect score and 0 represents the least skill (in terms of SR).

3.3.3 Receiver Operating Characteristic curves

Receiver operating characteristic (ROC) curves measure a probabilistic nowcast's ability to discriminate between convective events that happened and convective events that did not (the higher the area under a ROC curve, the more efficient it is at this). Each nowcast is thresholded at 11 equally spaced values (to provide a high enough density of points) within a given probability range. This probability range will vary depending on the range of probabilities produced by the nowcast. For example, a nowcast that produces probabilities between 0 and 1 will be sampled in the range [0,1] in 0.1 increments. The binary outputs are then compared to the corresponding binary observation and the POD and POFD scores are calculated (resulting in 11 scores – one for each threshold). This process is applied to the entire test set and the mean POD is plotted against the mean POFD at each threshold. Finally, this is repeated for j nowcast lead times, which produces j POD – POFD lines that span the probability range. A

skilful nowcast will maximise the POD and minimise the POFD, which results in a high Area Under the Curve (AUC) score – the key metric. ROC curves are used for evaluation in chapters 4 and 5.

3.3.4 Reliability diagrams

Reliability diagrams give insight into how a probabilistic prediction of an event compares to the observed frequency of that event. The nowcast probabilities are binned to create 10 sub-group bins and, for each sub-group bin, the frequency of observed events is calculated. The mean probability within each sub-group bin is then plotted against the corresponding frequency of observed events. For a perfectly reliable nowcast the predicted probability of events will exactly match the observed frequency of events, resulting in a line along the x=y line. Reliability diagrams are used for evaluation in chapters 4 and 5.

Chapter 4

Evaluating pySTEPS optical flow algorithms for convection nowcasting over the Maritime Continent using satellite data

4.1 Introduction

As discussed in chapter 2, the MC experiences regular convective activity that rapidly develops to form impactful storms. Radar data in the region is unreliable due to connection issues and blocking mountainous terrain. Instead, satellite-based nowcasting tools can be used to provide continuous coverage; however, there has been little focus on evaluating these tools across the MC.

pySTEPS (Pulkkinen et al. 2019) is a free, open-source Python library that provides modules for a variety of optical flow-based nowcasting methods (see section 2.5.1 for optical flow description). The library is designed for use on radar data and has been used to show skilful prediction of stratiform precipitation in the mid-latitudes (Han et al. 2022; Imhoff et al. 2020). To the best of the authors' knowledge, the only study that has applied pySTEPS to satellite data over the Tropics is Burton et al. (2022), who produced nowcast skill up to a 3 hour lead time (on a 64 km scale) over West Africa. It is this result that motivates the application of pySTEPS to the

MC.

This chapter presents the evaluation of both deterministic (LK) and probabilistic (STEPS) nowcasting algorithms by applying pySTEPS to satellite data over the MC. The aim is to highlight their strengths and weaknesses and demonstrate their potential use as an operational nowcasting system.

4.2 Data and methods

4.2.1 Data

BT data from the Himawari-8/9 satellites (see section 3.1 for satellite details) is used as input to the nowcasting algorithms. The data are selected from the DJF season, which is the peak season for convection over the MC (Birch et al. 2016), for five seasons from 2015/2016–2019/2020. In line with previous literature, a 235 K threshold was applied to the BT images to try to include convective regions only (Crook et al. 2024; Feng et al. 2021; Machado and Laurent 2004; Roca et al. (2017)).

Figure 4.1 shows the climatological diurnal cycle of minimum BT across the domain. Across the majority of islands in the MC, minimum BT is observed during the afternoon and evening periods. This is typically the most convectively active period over land (Yang and Slingo 2001), when the initiation and development of storms generate the coldest cloud top temperatures. As these storms reach maturity overnight, they often begin to propagate offshore (in land propagation is also observed over Borneo), resulting in minimum BT occurrence during the early morning period near the coast and over the sea. The timing of minimum BT over the sea in the northeast of the domain produces a less coherent pattern. This may be a reflection of the stochasticity of convection over a region where the meteorology is likely to be less influenced by the land (due to fewer surrounding islands).

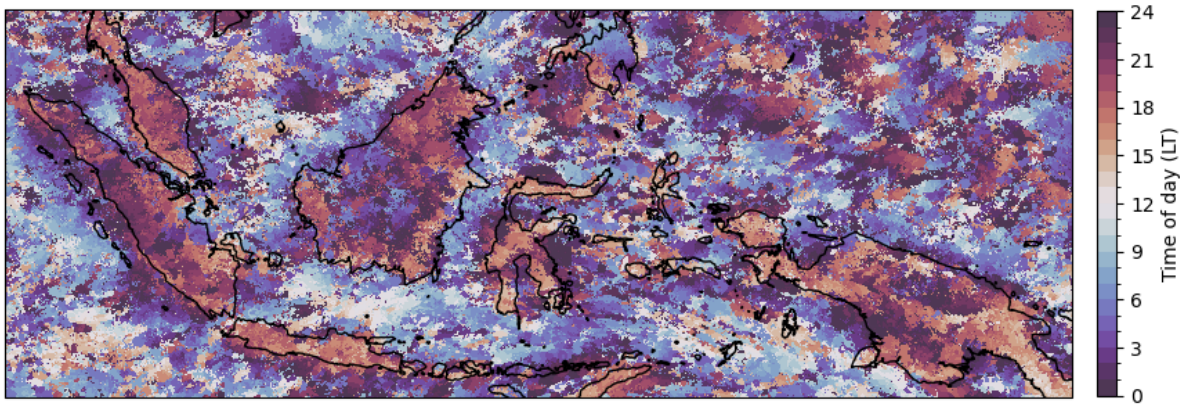


Figure 4.1: Mean diurnal cycle of minimum brightness temperature across the MC domain for the five DJF seasons from 2015/2016 - 2019/2020.

Nowcasts were produced using three BT maps as input, spaced evenly apart by 1 hour, starting with the latest observation. A total of 3,496 nowcasts were produced for initialisation times every 3 hours from 00:00 LT to 21:00 LT to incorporate the diurnal variability of weather over the MC, with the number of nowcasts at each initialisation time shown in Table 4.1. In order to avoid issues of new convection entering the edge of the domain, which cannot be reproduced (optical flow can only propagate convection that exists in the domain at the nowcast initialisation time), the nowcasts were first produced using BT data on a 15°S–15°N, 90°E–153°E domain and then evaluated on a 10°S–10°N, 94°E–149°E domain, which still includes the major islands of the MC (Figure 4.2).

Initialisation time (LT)	**Number of nowcasts**
0000	441
0300	442
0600	422
0900	441
1200	440
1500	429
1800	441
2100	440

Table 4.1: The number of nowcasts that were produced at each initialisation time throughout the day.

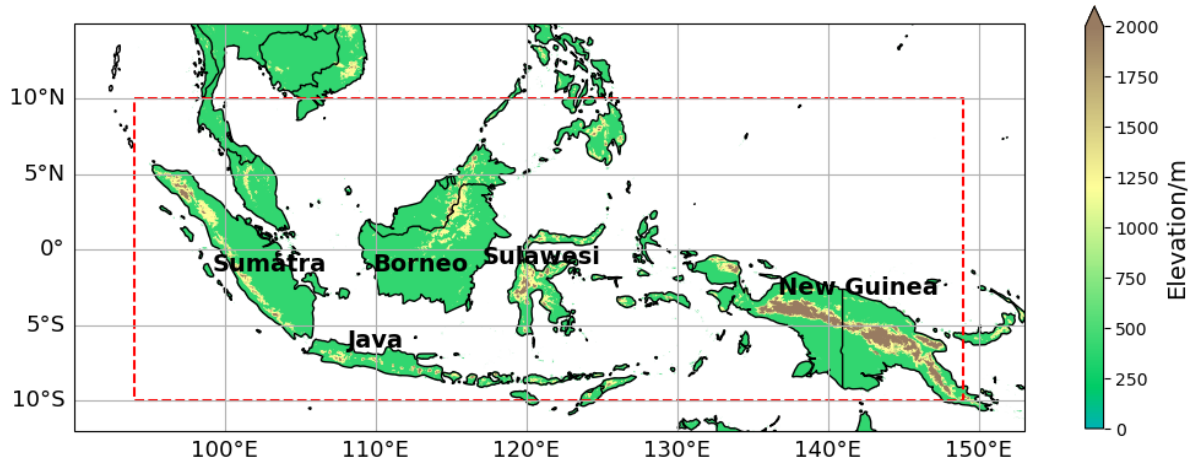


Figure 4.2: Orographic map of the Maritime Continent showing the domain over which the nowcasts were generated and the domain over which they were evaluated (dotted red line).

4.2.2 pySTEPS optical flow algorithms

The LK algorithm and STEPS algorithm (see section 3.2 for full descriptions of both) are used to generate the deterministic and probabilistic BT nowcasts. In this study, three consecutive BT images (the current BT observation and the two images prior) are inputted into the algorithms. For advecting the BT fields along the motion vectors produced by each algorithm, pySTEPS implements the backward-in-time semi-Lagrangian advection scheme (Germann and Zawadzki 2002).

This study uses a persistence nowcast (see section 3.3 for persistence of description) as a comparison baseline model for the optical flow algorithms.

4.2.3 Fractional skill score methods

The stochastic nature of convection in the MC makes it extremely challenging to nowcast the precise location (pixel-to-pixel) of convective activity. When evaluating a nowcast's ability to predict convection on a pixel-to-pixel basis, the nowcast may be broadly correct but slightly misaligned in location. If we simply take the difference between the nowcast and the verification, this leads to the double-penalty problem: firstly, the model is penalised for a miss, and secondly it is penalised for a false alarm in the slightly misaligned location. To overcome this problem Roberts and Lean (2008) developed a method known as the fractional skill score (FSS), which enables a forecast to be verified on a range of spatial scales as opposed to a pixel-by-pixel basis,

allowing leeway for minor misalignment. The FSS firstly creates two binary fields from the nowcast field and the observation field by using the 235 K threshold – any pixel with a value below this is set to 1, and any pixel with a value above this is set to 0. An n×n kernel is then convolved with both binary fields, where n is the desired spatial scale set by the user, and the fraction of pixels within the kernel that have a value of 1 is calculated. The mean squared error (MSE) between the fraction of 1's in the observation kernel, $O_{(n)}$, and the fraction of 1's in the nowcast kernel, $M_{(n)}$, is then calculated:

$$MSE_{(n)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[O_{(n)i,j} - M_{(n)i,j} \right]^2, \tag{4.1}$$

where N_x and N_y are the number of pixels in the longitude and latitude direction, respectively. Because $MSE_{(n)}$ is highly dependent upon the frequency of the event, it must be compared to the MSE of a relatively low-skill reference nowcast in order to provide any usefulness, defined in Murphy and Epstein (1989) as:

$$MSE_{(n)\text{ref}} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[O^2_{(n)i,j} + M^2_{(n)i,j} \right]. \tag{4.2}$$

The final FSS is then calculated as:

$$FSS = 1 - \frac{\text{MSE}_{(n)}}{\text{MSE}_{(n)\text{ref}}}. \tag{4.3}$$

The nowcast can be evaluated at different spatial scales by changing the value of n. In this study 10 km, 20 km, 60 km, 100 km and 200 km were chosen as the spatial scales. This range of scales allows a nowcast to be evaluated in its ability to predict convection on a range of scales. An FSS of 1 can be interpreted as a perfect score whereas an FSS of 0 can be interpreted as a nowcast with no skill. A threshold value for FSS above which a nowcast is useful is given by:

$$FSS_{\text{useful}} \geq 0.5 + \frac{f}{2}, \tag{4.4}$$

where f is the fractional coverage of pixels with a value of 1 over the entire domain. As f becomes small then $\text{FSS}_{(useful)}$ can be approximated by:

$$FSS_{\text{useful}} \geq 0.5. \tag{4.5}$$

This is the basic approach to the FSS, which calculates a single score for each nowcast to describe the skill over the whole domain. However, often the skill of a nowcast will vary across the domain due to differences in environments (e.g. land, sea and mountains) and the different interactions that result from these changing environments. Woodhams et al. (2018) developed an adapted version of FSS known as the localised fractional skill score (LFSS), which enables the skill of a nowcast to be evaluated at each pixel across the domain, resulting in a spatial map of FSSs. The LFSS is calculated by adapting equation 4.1 and 4.2 so that instead of dividing the sum of the squared error between $O_{(n)}$ and $M_{(n)}$ over the spatial domain, it is divided over the time domain (i.e. the number of time steps). This results in replacing equation 4.1 and 4.2 with,

$$MSE_{(n)} = \frac{1}{N_t} \sum_{t=1}^{N_t} \left[O_{(n)t} - M_{(n)t} \right]^2, \tag{4.6}$$

and,

$$MSE_{(n)\,\text{ref}} = \frac{1}{N_t} \sum_{t=1}^{N_t} \left[O_{(n)t}^2 + M_{(n)t}^2 \right], \tag{4.7}$$

where N_t is the number of time steps.

It is important to consider the effects of bias when interpreting the FSS score. When applying a threshold to a biased nowcast (e.g. systematically too cold BT values), the fraction of grid cells exceeding the threshold will differ between the observation and the nowcast. This means that, even on large spatial scales, the FSS score can never reach a value of 1. This can result in the nowcast appearing to have lower skill than it really does in terms of spatial scale.

4.3 Results

4.3.1 Deterministic nowcasting - Lucas-Kanade algorithm

Figure 4.3e-g show an example of a nowcast (each lead time is produced using the same T-0, T-1 and T-2 observations) produced by the LK algorithm for a qualitative assessment of the skill

against observations (Figures 4.3a-d), whilst Figures 4.3h-j provides the LFSS (evaluated on a 20 km scale to show clearly defined differences in skill) at each time step of the nowcast (where N_t=1) for a quantitative assessment. This particular set of observations contains convection on a range of scales, with regions of propagation and regions of initiation, providing a good example to evaluate the LK algorithm on a range of capabilities (for this reason the same example is used throughout the study). At T-0, the observed organised, large-scale convection (e.g. north of Borneo) approximately maintains its shape through T+1 and T+3 and then starts to change in structure at T+6 (e.g. east of Sumatra). There is also the development of relatively smaller-scale convection observed during the T+3 and T+6 hours lead times (e.g. over New Guinea).

Visually, the deterministic LK approach appears to predict the propagation of large-scale, organised convection well. The T+1 nowcast best resembles the corresponding observation due to the least amount of new convection developing during this time, as well as little propagation of the organised convection (the nowcast at T+1 closely matches the persistence nowcast). This is seen in Figure 4.3h, which shows high skill in the organised convective regions over the majority of the domain. As the lead time increases, westward propagation of the large-scale convection is observed, which appears to be effectively tracked by the LK nowcast at T+3 (e.g. east of Sumatra). This is confirmed in Figure 4.3i, which shows the skill of the nowcast at T+3 remaining high over the regions of organised convection. There is, however, a clear increase in areas of low skill at T+3 and T+6 due to the LK nowcast being unable to reproduce the initiation-growth-decay (IGD) of convection. At T+6, the change in structure of convection also contributes to the majority of the domain experiencing low skill. The nowcast at this lead time shows the least resemblance to the observations with only the largest regions of predicted convection providing any skill (Figure 4.3j).

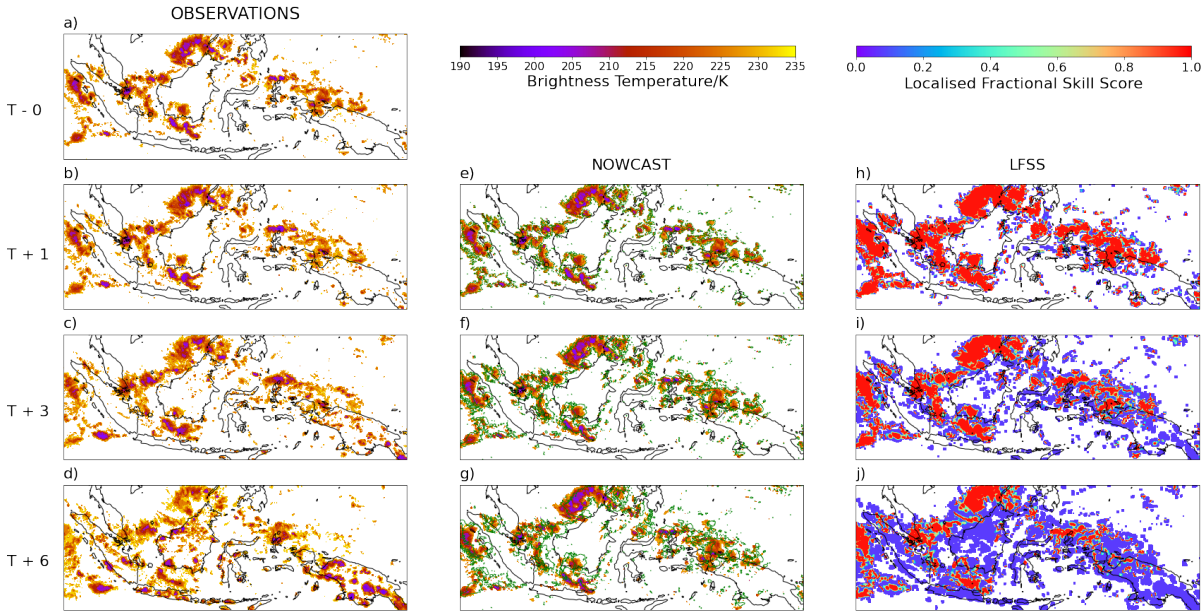


Figure 4.3: Panels (a–d) are maps of BT (<235 K) observations showing convection propagating across the MC on 11 December 2019 starting from the initial observation at 09:00 LT (T - 0), followed by the observation at 1 hour (T + 1), 3 hours (T + 3) and 6 hours (T + 6) later. Panels (e–g) are the nowcasts produced by the LK algorithm at each of the corresponding observations, and the green lines show the contour of the corresponding persistence nowcast. Panels (h–j) are the LFSS maps produced by evaluating each nowcast against the corresponding observation on a 20 km spatial scale.

Figure 4.3 highlights some of the key advantages and disadvantages of the LK algorithm. Overall, it does well at predicting the propagation of large-scale, organised convection. However, because of the principle that underlies optical flow – that each pixel maintains its intensity between time steps – it is unable to capture the IGD of convection. Smaller-scale convection exhibits higher rates of change in its evolution (Venugopal et al., 1999) and so has the greatest uncertainty associated with it. Initially, this explains why the majority of low skill is seen at smaller scales (Figure 4.3h-i). However, at T+6 the difference in small-scale features between the observations and the nowcast becomes more widespread, and so the low skill spreads further across the domain (Figure 4.3j).

Figure 4.4 shows the mean FSSs for all 3,496 LK nowcasts (Table 4.1) and their corresponding persistence nowcasts, evaluated at spatial scales of 10 km, 20 km, 60 km, 100 km and 200 km. The 10 km spatial scale is the smallest scale of evaluation; hence it consistently produces the lowest scores. However, the model still shows good skill on this scale (FSS$\geq$0.5) at a lead time of 4 hours. Doubling the spatial scale to 20 km increases the skilful lead time by $\sim$1 hour. At the 60, 100 and 200 km spatial scales, the LK nowcasts show skill across all lead times with

FSS scores of ∼0.54, ∼0.61 and ∼0.75, respectively, at the 6 hour lead time.

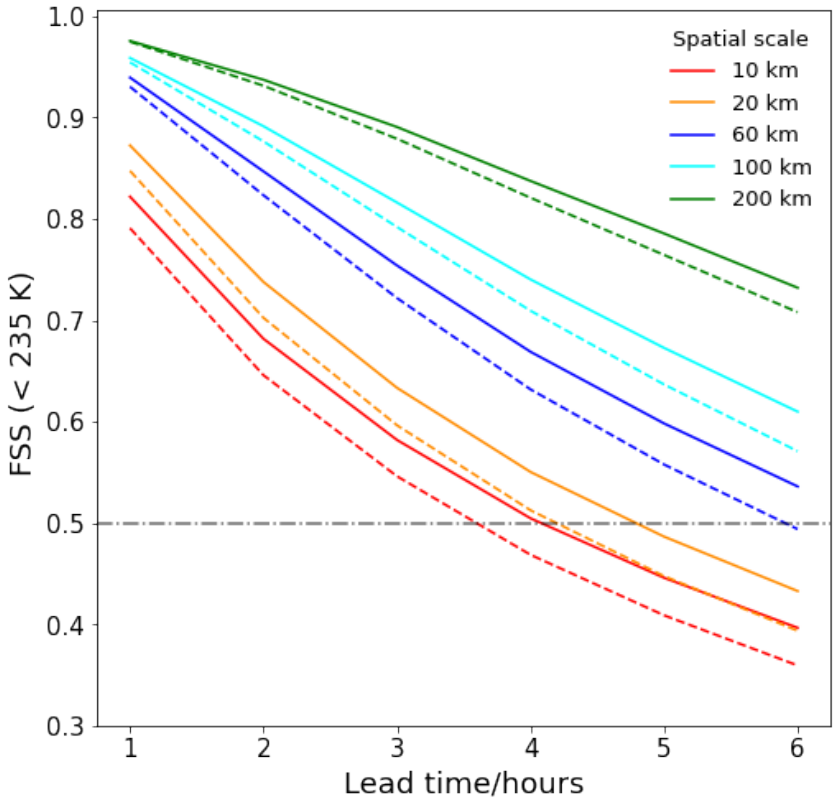


Figure 4.4: Composite FSSs against lead time for 3,496 LK nowcasts (solid line) and persistence nowcasts (dashed line), evaluated at a threshold of 235 K for 10 km, 20 km, 60 km, 100 km and 200 km spatial scales. The grey horizontal line marks the 0.5 FSS line, which is considered the cut-off for nowcast skill.

Skill reduces with lead time for all spatial scales and increases with spatial scale at all lead times. On average, the LK nowcasts outperform the persistence nowcasts at all lead times and for all spatial scales. Both the persistence and the LK nowcasts maintain the same structure and intensity at each lead time (hence relatively smaller skill difference at 1–2 hour lead time). However, the LK nowcasts propagate the convection across the domain, whereas the persistence nowcasts remain stationary. This explains the increasing added value of the LK nowcast with lead time, as the observations move further from the persistence nowcast and the skill difference increases. Greater added value of the LK nowcast over persistence is also seen at smaller spatial scales. For example, the skill gap between the persistence nowcasts and the LK nowcasts at the 6 hour lead time is greater for 10 km spatial scale compared to 200 km spatial scale. The trend of FSSs across the shown lead times also varies for different spatial scales. At smaller

spatial scales the nowcast is being evaluated on its ability to predict smaller-scale convection, which changes most rapidly/unpredictably, resulting in a higher rate of decrease in skill. The FSS evaluation on higher spatial scales smooths out these smaller-scale convection changes, and so a much lower rate of decrease in skill is seen. The rate of decrease of the FSS evaluated on the 10 km spatial scale appears to show its largest rate of skill decrease at 1–2 hour and then decelerates over the following time intervals. On the other hand, the rate of skill decrease for the 200 km spatial scale increases with lead time.

Figure 4.5 shows the mean LFSS for 3 hour lead time LK nowcasts over the MC, evaluated at a 100 km spatial scale. Evaluation on this scale has been used as it is able to clearly highlight the variations of skill across the domain. For all nowcast initialisation times there is consistent noise in the skill over the sea. This may be representative of the stochastic nature of convection initiation over the sea. It therefore would be expected that, if the period of evaluation were extended beyond 2015–2020 (i.e. increasing the number of nowcasted events used in the averaging over time in equations 4.6 and 4.7), the LFSS noise field would become smoother over the sea. During the overnight and morning initialisation times (Figure 4.5a–d) there is, on average, high skill over the majority of the domain (Figure 4.5i). This is to be expected as at these times the majority of convection has formed large-scale, organised cloud systems with very few small-scale convection initiations occurring. These larger-scale systems will most often be propagating offshore of the islands (as part of the diurnal cycle), which the LK algorithm is most effective at capturing. Furthermore, overnight and during the morning there is, on average, relatively less convective activity than during the day meaning that reduced skill from inaccurate propagation predictions is minimised.

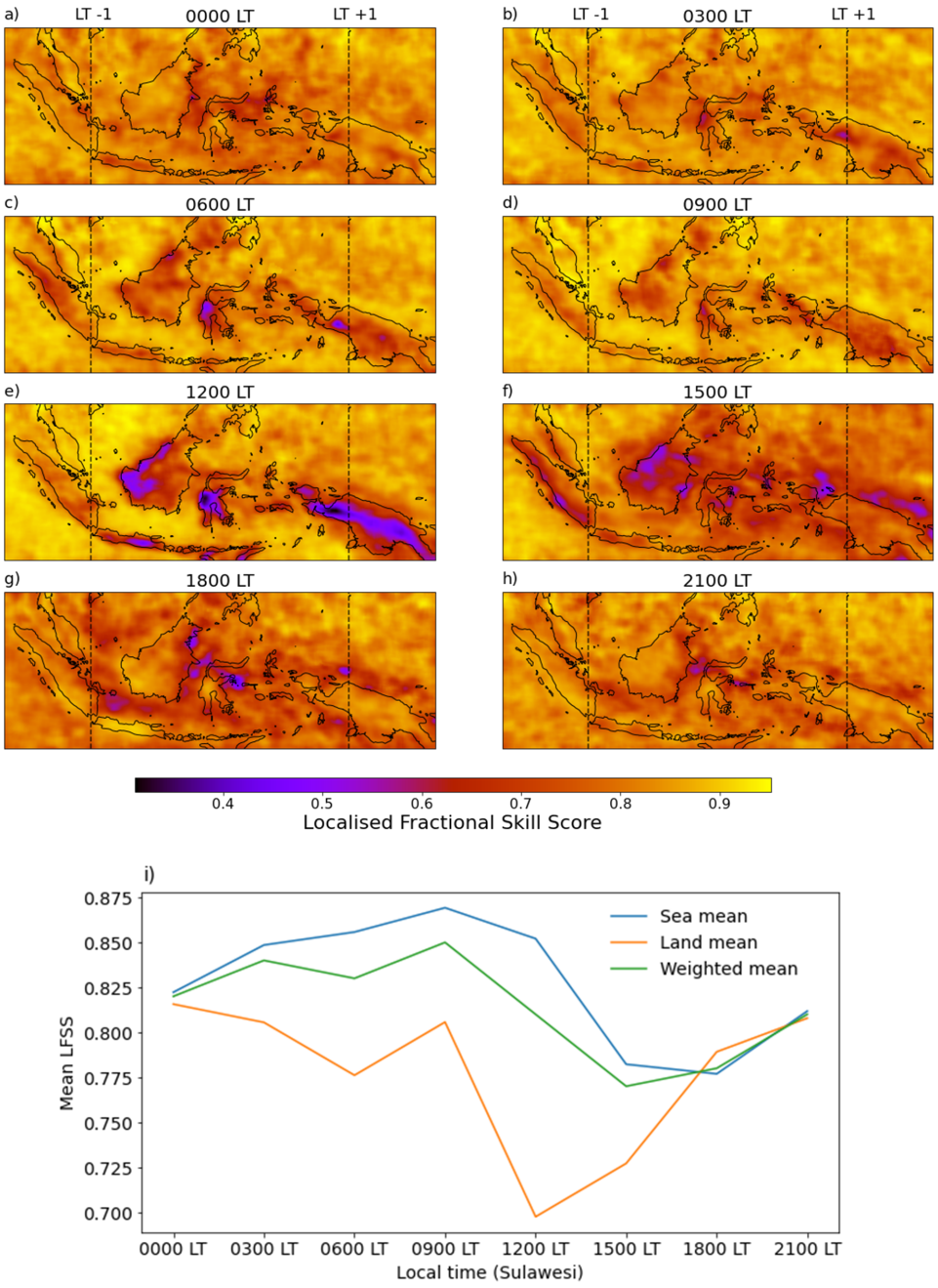


Figure 4.5: (a–g) Composite LFSS maps for 3 hour lead time LK nowcasts that were initialised at 00:00 LT, 03:00 LT, 09:00 LT, 12:00 LT, 15:00 LT, 18:00 LT and 21:00 LT, respectively. The LFSS was evaluated on a 100 km scale, and the local time is with reference to Sulawesi (the vertical dotted lines show the time difference across the domain). (i) The mean LFSSs at each initialisation time for sea only, land only and the whole domain.

Between 09:00 LT and 12:00 LT there is a significant drop in the mean skill over land (Figure 4.5i). Figure 4.5e–f show these distinct regions of low skill over land, which are tightly constrained to the coastal and mountainous regions of the islands and are closely tied to the diurnal cycle of convection over the MC. Convection begins to initiate and develop over the coastal and

mountainous regions of the islands in the early afternoon, which is not present at 12:00 LT. The LK nowcast is unable to capture this new convection, resulting in low overland skill for the 12:00 LT initialisation. Over these locations at this time of day, the LK algorithm would not be a skilful nowcasting tool. In West Africa, Burton et al. (2022) also showed a similar drop in skill due to the development and growth of new convective cells. In the MC, the low skill is still seen at 15:00 LT but to a lesser extent. At this initialisation time the T-0 observations that are inputted into the LK algorithm will, on average, contain the majority of the convection that has initiated over the early afternoon. This convection will likely remain stationary over this time but will be growing in size. Therefore, the persisting low skill at 15:00 LT is representative of the LK nowcasts' inability to predict the growth of convection. Furthermore, at 15:00 LT, a region of low LK skill is observed over the eastern part of the domain. This likely represents the algorithm's inability to capture the full BT field produced by the propagation of storms over this region.

The development of convection starts to slow as storms reach their mature stage in the evening. Less growth results in higher skill over the land. At 21:00 LT the LFSS map looks similar to the overnight LFSS maps with high skill across the entire domain. Over this 3 hour nowcast period the LK algorithm has shown good skill at being able to nowcast the propagation of mesoscale convective systems offshore, which developed overland during the afternoon.

Understanding that the LK algorithm is unable to predict the IGD of convection means that, by identifying anomalous regions of low skill, LFSS maps can be a useful tool for identifying local effects due to land–sea interactions. An example of this is seen in a region of low skill over the northeast coast of Borneo at 18:00 LT. On average, at this time of the day a land breeze begins to develop along the entire concave-shaped coastline, potentially causing convergence near the middle. This convergence may then lead to the initiation of convection that the LK algorithm is unable to predict.

To provide a baseline comparison for the LK algorithm's spatial skill, Figure 4.6 presents the mean LFSS for 3 hour lead time persistence nowcasts, evaluated throughout the day at a 100 km spatial scale. Figure 4.6i shows that, overall, the LK algorithm has greater spatial skill than persistence over the land and sea (Figure 4.5i), at each evaluation time. The distribution of spatial skill for persistence also shows similar trends to the LK algorithm. Relatively high skill is observed across the domain during the night and morning period (Figure 4.6a-d), and

lowest skill is observed over the land at midday and during the afternoon (Figure 4.6e-f). The greater drop in persistence skill at 12:00 LT (Figure 4.6i) compared to LK can be explained by a greater spread of low skill regions over New Guinea, Java, Sulawesi and Borneo (Figure 4.6e). As with the LK algorithm, the persistence spatial skill increases again during the evening (Figure 4.6g-h), when convective development activity reduces.

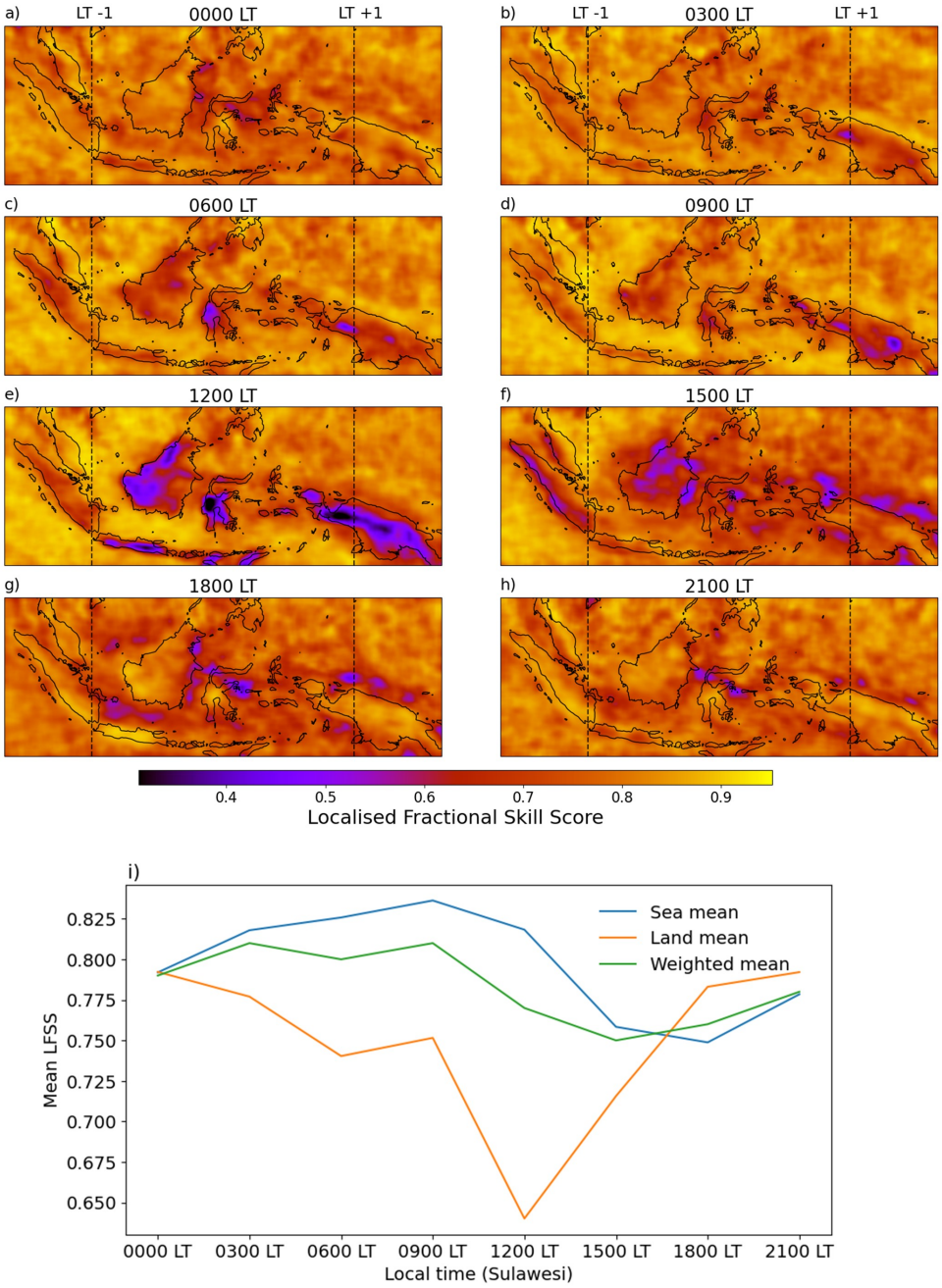


Figure 4.6: Same as Figure 4.5 but for 3 hour lead time persistence nowcasts.

4.3.2 Ensemble nowcasting - STEPS algorithm

Figure 4.6a provides an example of a 20-member ensemble nowcast (generated using the same T-2, T-1 and T-0 set as for the LK algorithm example in Figure 4.3), with a lead time of 3 hours, produced by pySTEPS's implementation of STEPS. Over the 3 hour period the main differences between the T-0 and T+3 observations are around New Guinea where new convection develops over the land and the convection north of the island becomes more scattered (Figure 4.6b-c). Visually, each ensemble member provides a good prediction of the large-scale convection (e.g. north of Borneo), with little difference between the members in the predicted shape and structure. The main differences between the ensemble members come from differences in the stochastic fields injected for each member. This is seen in the varying levels of BT intensities in the large-scale convection in each member. For example, the BT intensity over northern Borneo in ensemble member 13 is greater than in ensemble member 2. Differences are also seen between each ensemble member in the distribution of the predicted small-scale convection (e.g. over the Philippine Sea). Over Borneo, some members have predicted small-scale convection which approximately aligns with the new convection observed at T+3 (e.g. ensemble member 18), whereas some members have predicted no convection here at all (e.g. ensemble member 17).

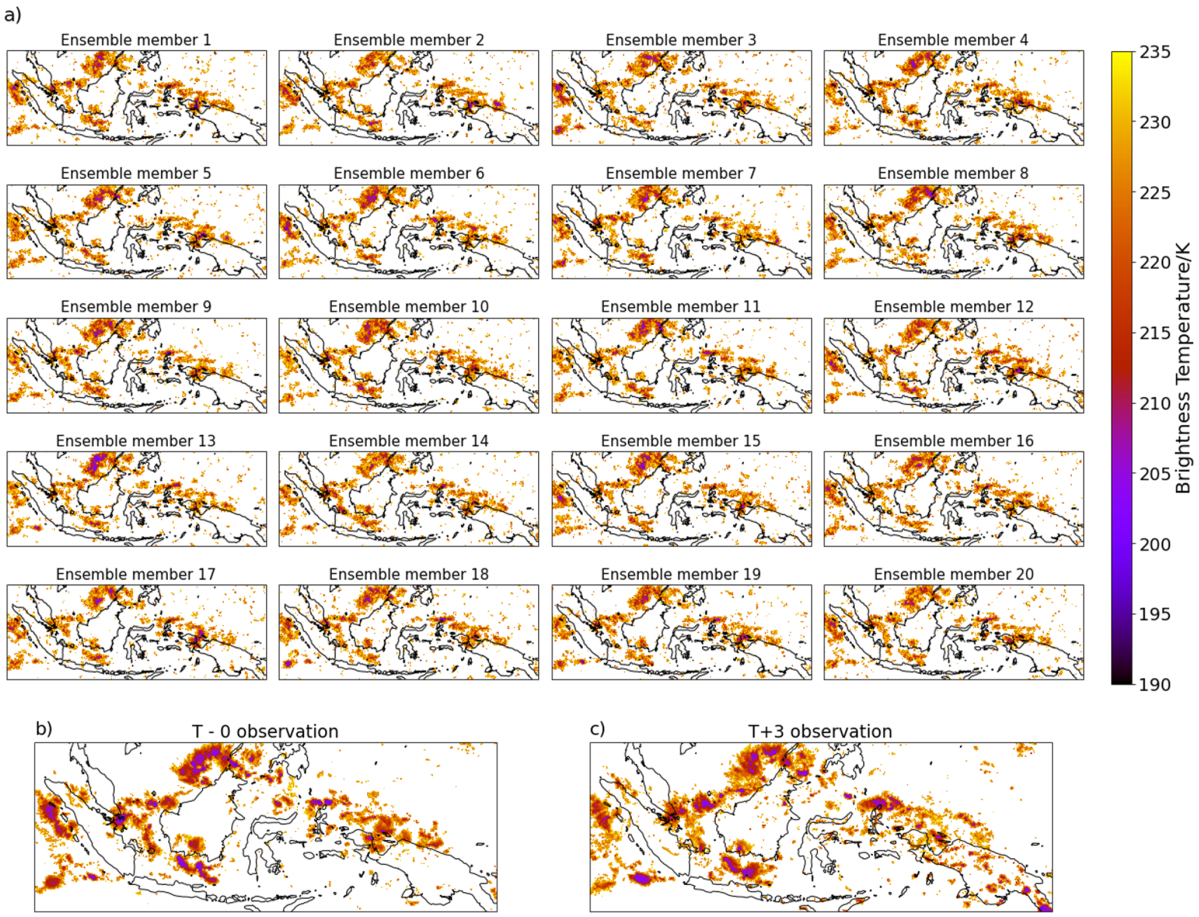


Figure 4.7: a) An example of a 20-member ensemble, 3 hour lead time BT (<235 K) nowcast produced by the STEPS implementation of pySTEPS on 11 December 2019. (b) The observation at the nowcast initialisation time and c) the observation 3 hours later

One of the key aims of STEPS is to address the uncertainty in the evolution of small-scale convection. In all members the convection northwest of New Guinea appears much noisier than in the T-0 observation. STEPS has recognised this as a region of uncertainty and addressed it by injecting noise at this scale. When compared to the T+3 observation, it can be seen that the convection does in fact become more scattered and dissipated, and so, although STEPS has not been able to precisely predict the new shape of the scattered convection, it has been able to capture the unpredictable nature of the evolution of this small-scale convection.

The 20-member ensemble in Figure 4.7 has been used to produce the probabilistic nowcast in Figure 4.8 (extended to 1, 3 and 6 hours lead times). This probabilistic nowcast uses a threshold of 235 K, therefore including all the pixels that were used to produce each ensemble nowcast. At T+1 the probabilistic nowcast shows a high degree of certainty in its prediction of the shape and location of convection, meaning that there is little variance between ensemble

members. The T+1 lead time is the first time step prediction that the algorithm makes, and so it contains the least amount of stochastic noise in the extrapolation motion field, hence the least amount of member variance. As the lead time increases, more stochastic noise is injected into the extrapolation motion fields, and so the uncertainty of the probabilistic nowcast increases. This can be seen in the reduction in high probabilities over the regions enclosed within the 20% contour. A reduction of high uncertainties within the 20% contour is also matched with a greater spread of low ensemble probability across the entire domain (outside the 20% contour) at T+3 and T+6.

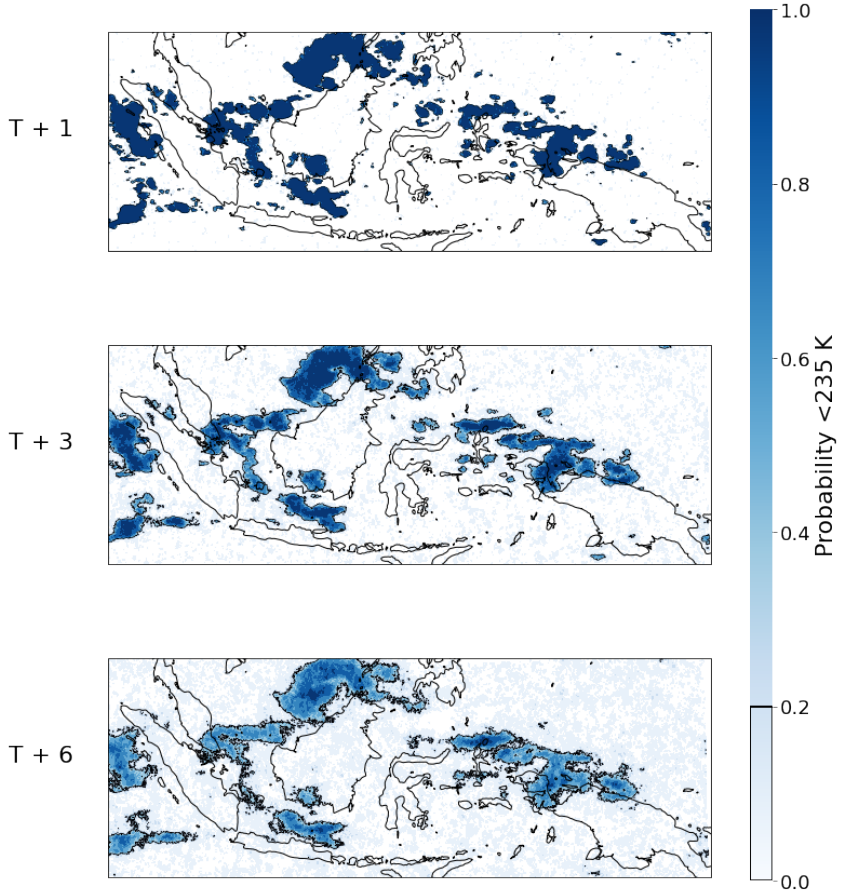


Figure 4.8: An example of a probabilistic BT nowcast for lead times of 1, 3 and 6 hours produced by the STEPS implementation of pySTEPS on 11 December 2019. A BT threshold of <235 K is used in order to include all pixels in the nowcast.

Figure 4.8 also shows the number of small-scale features in the probabilistic nowcast reducing at longer lead times. For example, at T+1 the convection northwest of New Guinea appears scattered in small blobs, whereas at T+6 STEPS has smoothed out this small-scale convection into a larger region. Again, this is evidence of the algorithm's attempt to address the uncertainty in the evolution of small-scale convection by replacing it with stochastic noise.

Figure 4.9a and b show the mean ROC and reliability curves (see section 3.3 for details on these verification methods) for STEPS nowcasts initialised in the morning (09:00 LT) and afternoon (15:00 LT), evaluated over the sea and the land. For both surface types and initialisation times, the POD increases at each threshold meaning that, even as more uncertain regions enter the evaluation, the proportion of hits to misses increases. Furthermore, the POD exceeds the POFD at each threshold indicating that STEPS has skill in predicting regions <235 K BT over the MC. The greatest POD–POFD difference, which is considered the optimum threshold for a probabilistic nowcast, is shown at the ≥10% likelihood threshold for all of the curves.

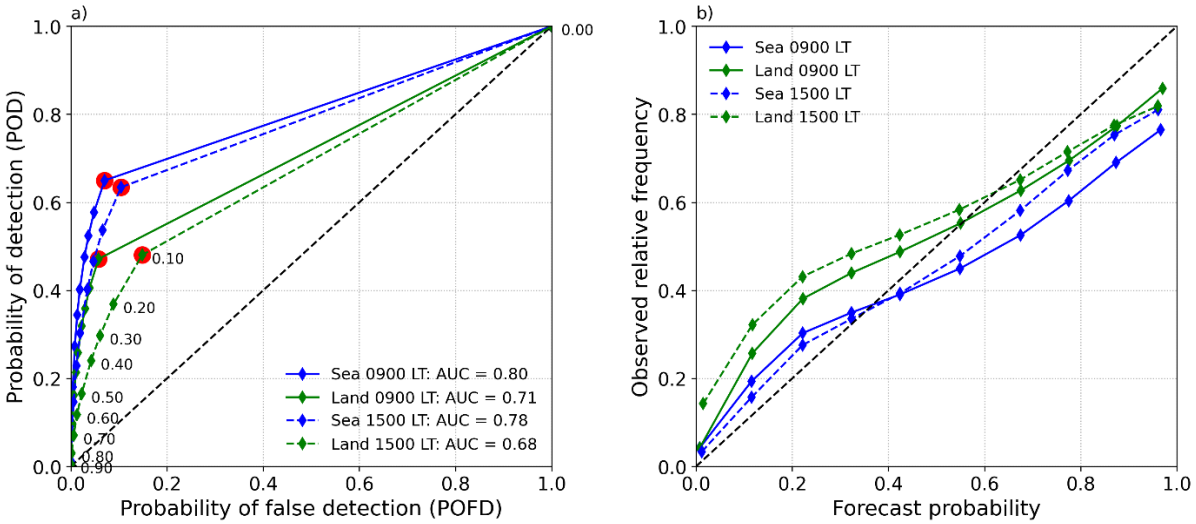


Figure 4.9: Composite (a) ROC and (b) reliability curves over the sea and land for STEPS 3 hour lead time nowcasts, initialised at 09:00 LT (441 nowcasts) and 15:00 LT (405 nowcasts), with a threshold of <235 K. The numbers next to the green points in (a) represent the thresholds used to evaluate the nowcast at different likelihoods. Red dot represents the optimum threshold.

For both initialisation times in Figure 4.9a, STEPS produces higher area under the curve (AUC) scores over the sea (0.80 and 0.78 for 09:00 LT and 15:00 LT, respectively) than over the land (0.71 and 0.68 for 09:00 LT and 15:00 LT, respectively), meaning that STEPS has more skill over the sea at these times. This can be explained by lower POD scores over the land, which is due to the STEPS algorithm being unable to capture the new convection that most often develops there (increasing the number of misses). Furthermore, a comparison of each region within initialisation times suggests that STEPS is slightly more skilful in the morning (average AUC of 0.76) than in the afternoon (average AUC of 0.73). The morning–afternoon difference in POFD is also much greater over the land than the sea. This is likely due to a greater decrease

in correct negative scores over the land, caused by new convection initiating where there was previously none, which STEPS is unable to predict.

The positive slope between all the points for each reliability curve in Figure 4.9b means that, over both surface types, as the observed frequency of events increases, STEPS predicts a higher likelihood of that event occurring. This shows that overall STEPS is able to produce a reliable nowcast for these initialisation times. However, for both regions and times of day, STEPS presents a lower probability than the observed frequency for probabilities $\leq$0.35 and a higher probability than the observed frequency for probabilities $\geq$0.65. The distribution of ensemble predictions is therefore under-dispersive, meaning that the spread of predictions falls within the spread of observations, i.e. it does not provide an optimum estimate of uncertainty.

When comparing the two surface types it can be seen that the under-prediction at lower nowcast probabilities is greater over the land, meaning the ensemble distribution has less variance and STEPS is better at capturing uncertainty over the sea for these initialisation times. However, at higher nowcast probabilities, more over-prediction is seen over the sea compared to over the land, meaning that STEPS becomes too confident at predicting higher-likelihood events over the sea.

The under-dispersive feature of STEPS over the MC is due to low ensemble member variance, which (as previously mentioned) can be exemplified by visually assessing the lack of diversity between the ensemble members in Figure 4.7. The main source of ensemble member variance comes from the differences in the stochastic noise fields that STEPS injects into the nowcasts, and so, increasing the range of noise field intensities, or simply adding more members, would likely help to reduce this under-dispersive feature.

4.3.3 Comparison of STEPS, LK and persistence

By applying a threshold to a probabilistic STEPS nowcast, it is possible to produce a deterministic STEPS nowcast (in the form of a binary field) that can be evaluated using the FSS and directly compared to the corresponding LK and persistence nowcast. Figure 4.10 shows the mean FSSs for 3,496 LK nowcasts (solid), persistence nowcasts (dashed) and STEPS deterministic nowcasts produced using a threshold of $\geq$10% (dotted; STEPS10). The choice of threshold was based on the results of Figure 4.9, which show that, for both the morning and evening initialisation times, the optimum likelihood threshold was $\geq$10%.

At 10 km spatial scale, STEPS10 shows skill up to ∼5 hours and has the highest skill across all lead times, outperforming LK and persistence. At 60 km scales STEPS10 still outperforms persistence across all lead times but approximately equals the skill of LK from 1–3 hours' lead time. Onwards of 3 hours' lead time, STEPS10 shows higher skill than LK.

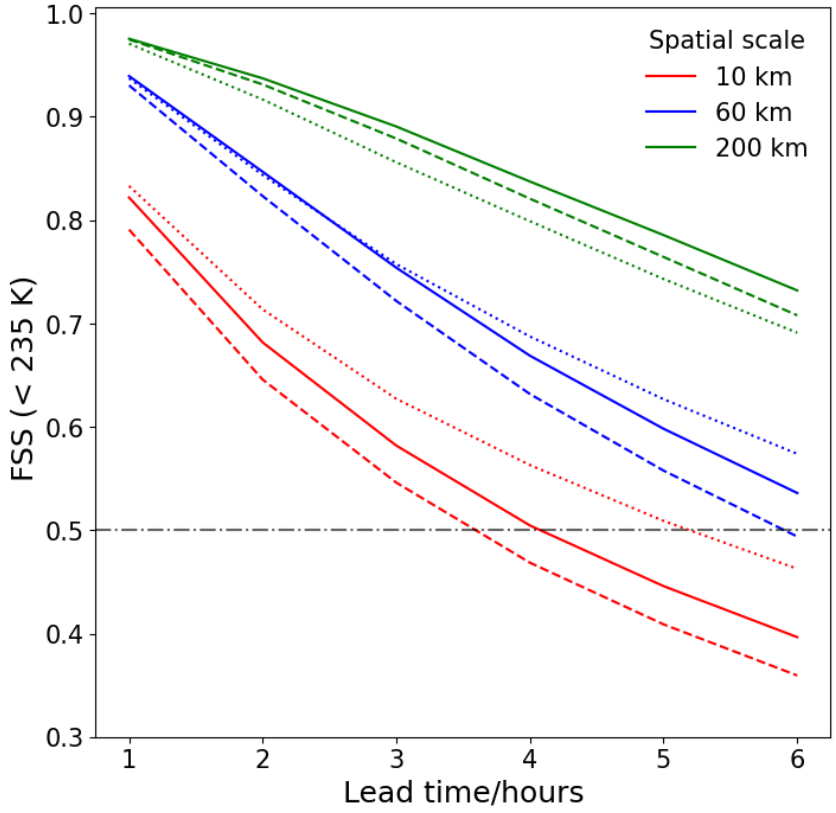


Figure 4.10: Composite FSSs against lead time for 3,496 LK (solid line), STEPS10 (dotted line) and persistence (dashed line) nowcasts, evaluated at a threshold of 235 K for 10 km, 60 km and 200 km spatial scales. The grey horizontal line marks the 0.5 FSS line, which is considered the cut-off for nowcast skill.

The added value of STEPS10 over both LK and persistence decreases between the 10 and 60 km spatial scales and, by the 200 km scale, STEPS10 shows the least skill out of all the nowcasts. This is likely due to less of the propagation being detected within the 200 km scale evaluation compared to the 10 and 60 km scales – hence LK tends towards persistence at greater scales. Unlike LK, STEPS changes the internal structure of the convection through the injection of stochastic noise. This change in the internal structure of convection (as opposed to change due to propagation) will be detected on the 200 km spatial scale and may contribute to a drop in performance relative to LK and persistence.

4.4 Conclusions and discussion

A deterministic (LK algorithm) and probabilistic (STEPS algorithm) implementation of the pySTEPS optical flow nowcasting library have been applied to satellite data over the MC to produce nowcasts with lead times of up to 6 hours.

Overall, the LK algorithm predicts the propagation of convection across the domain with skill (FSS $\geq$0.5) up to 4 hours on the 10 km scale (the smallest scale of evaluation) and up to at least 6 hours on the 60 km scale. Similarly, (Burton et al. 2022) used the LK algorithm to nowcast convective rain rates over West Africa. Although it is difficult to precisely compare the two sets of results (rainfall retrievals likely have more fine-scale variability and errors in rainfall retrievals will affect nowcast skill (Hill et al. 2020)) they are nevertheless comparable. Burton et al. (2022) show skill up to about 3 hours on the 64 km scale, whereas we would expect somewhat higher skill for BT nowcasts over the MC.

Similar to the findings in Burton et al. (2022), the LK algorithm was unable to predict the IGD of convection, which is a manifestation of the optical flow assumption – each pixel maintains its intensity across all time steps. This inability to predict IGD of convection is clearly seen when analysing maps of LFSS over the MC. Over the sea the LK algorithm shows, on average, good skill at all nowcast initialisation times due to convection being mostly propagating in nature. Over land, however, the model shows high skill in the morning and evening but much lower skill in the afternoon. In the early afternoon this is due to the initiation of convection, which is closely constrained to the mountains. Later in the afternoon the low skill is due to the growth of the convection that initiated earlier on and persists over the mountains. Over the mountainous regions of the MC during the afternoon, the LK algorithm would not be a useful nowcasting tool.

The STEPS algorithm aims to address the issue of unpredictability in the IGD of convection by injecting varying intensities of stochastic noise at different length scales to produce an ensemble nowcast. When analysing a probabilistic nowcast produced by a STEPS ensemble, it can be seen that the injection of noise has a smoothing effect, removing small-scale convection and maintaining the shape of the larger, more predictable convective regions.

A composite analysis of STEPS nowcasts was performed using ROC and reliability curves for 3 hour lead time predictions. Overall, the analysis showed that STEPS can produce both

skilful and reliable ensemble predictions for 3 hour lead times. When comparing land and sea at different times of day, it was shown that STEPS has highest skill over the sea during the morning (09:00 LT initialisation time) with an AUC score of 0.8 (compared to an AUC score of 0.71 over the land). Imhoff et al. (2020) also applied the pySTEPS implementation of STEPS to radar data over the Netherlands to produce nowcasts for a range of lead times. In their work they produced composite ROC curves for nowcasts with a lead time of 95–120 minutes, which had an AUC score of 0.81. The Netherlands experiences far less convective activity than the MC with the majority of its weather coming from propagating frontal clouds. This, therefore, further highlights the effectiveness of STEPS for the MC as it tries to predict convective clouds, which have a more unpredictable nature. However, the disadvantages of STEPS were revealed when analysing the reliability curves. A common feature across both times and regions was the under-dispersive ensemble distributions, which were more extreme over land. This highlights STEPS's inability to predict the low-likelihood events (e.g. new initiations) and capture the whole uncertainty of the observed system.

To compare STEPS with LK and persistence, a deterministic version of STEPS was produced by thresholding the probabilistic nowcasts at $\geq$10% (STEPS10). When evaluated, the STEPS10 nowcasts had higher skill at spatial scales of 10 km (across all lead times) and 60 km (from 3–6 hours). Therefore, not only does STEPS provide insight into the uncertainty of a convective system, but it can also derive a better single deterministic nowcast than LK and persistence at these scales. However, at a higher spatial scale of 200 km (where relatively less convection propagation is detected), LK nowcasts had the highest overall skill and the injection of stochastic noise produced by STEPS likely caused the STEPS10 nowcasts to have the lowest overall skill.

Continuous nowcasting over the entire MC is a requirement for EWSs, which does not currently exist. By providing both nowcast examples and composite nowcast analysis, this study has shown the effective application of a deterministic (LK) and probabilistic (STEPS) algorithm to satellite data, showing their potential to be used operationally over the whole of the MC. This work highlights the key strengths and weaknesses of both algorithms, providing important information to a potential forecaster using these tools.

Chapter 5

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

5.1 Introduction

Chapter 4 highlighted the capability gap of optical flow algorithms when nowcasting the IGD of convection. This study presents Simple Initiation and Intensification Nowcasting neural Network (SII-NowNet), an ML model that produces individual nowcasts of convection initiation and convection intensification in the Tropics, with lead times of 1 – 6 hours (in 1 hour increments).

Previous ML nowcasting studies in the Tropics (Han et al. 2019; Lagerquist et al. 2021; Lee et al. 2017) still require radar data for training, leaving a nowcasting capability gap in the 'radar-less' regions of the Tropics. SII-NowNet's simplicity derives from requiring only BT data for training and testing, meaning it can be applied to any other region of the Tropics.

The performance of SII-NowNet is first evaluated over Sumatra, Indonesia – one of the largest Tropical islands in terms of size and population, which experiences regular convective activity (Mori et al. 2004). SII-NowNet is evaluated against climatology and persistence baselines to identify a limit of usefulness, and the minimum number of training samples required before

it loses skill is quantified. A feature importance test is also applied to gain insight into SII-NowNet's decision-making. Finally, SII-NowNet is evaluated over Zambia to exemplify its applicability to other regions in the Tropics.

5.2 Data and methods

5.2.1 Data

SII-NowNet was developed and tested over the Sumatra, Indonesia domain (Figure 5.1a) using hourly BT data from the Himawari-8/9 satellites (see section 3.1 for satellite details) from 01/01/2018 to 31/12/2019 (17,327 images).

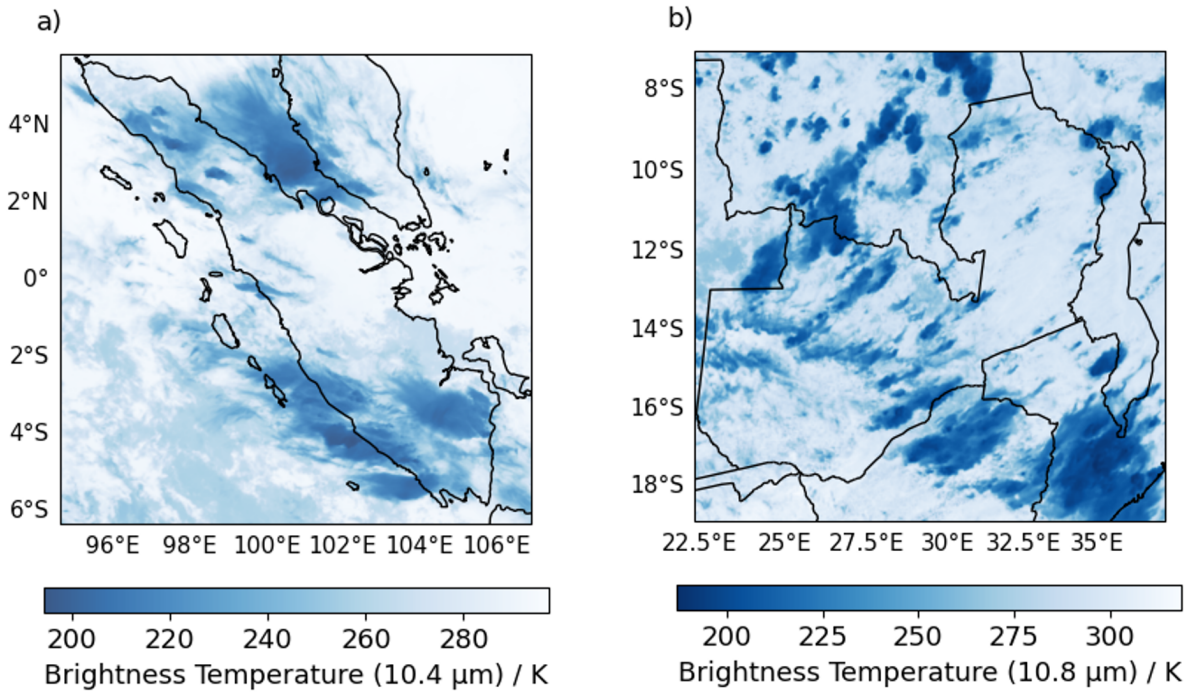


Figure 5.1: Examples of BT images over the two domains used in this study. a) Provides an example of a 10.4 µm BT image over Sumatra, Indonesia on 7th March 2019 at 20:00 LT. b) Provides an example of a 10.8 µm BT image over Zambia on 15th March 2024 at 15:00 LT.

For further testing and development of SII-NowNet over a new Tropical domain, Zambia is used as an exemplar region (Figure 5.1b). BT data from the Meteosat-10 satellite is used (see section 3.1 for satellite details). Although the Meteosat-10 BT wavelength (10.8 µm) is longer than the Himawari BT wavelength (10.4 µm), 10.8 µm is the closest available wavelength. As Zambia is used as an exemplar case study, a shorter period of data is used from 01/02/2024 to 01/03/2024

(4,459 images). Data from this period was used to evaluate two versions of SII-NowNet over Zambia: the original SII-NowNet (trained over Sumatra) and SII-NowNet-Zam (a new version with additional Zambia-specific training).

5.2.2 SII-NowNet nowcast setup

SII-NowNet requires a 416 x 416 x 2 pixel input matrix (Figure 5.2). The first two dimensions represent the spatial shape of the input (i.e. 2-D fields covering the target domains), and the third dimension represents the two predictor channels. The two predictor channels are:

1. BT T-0 hours
2. BT T-1 hours

where T represents the nowcast initialisation time. BT T-0 provides information on present atmospheric conditions, whereas BT T-1 provides information on conditions one hour before present. By learning the difference between BT T-0 and BT T-1, SII-NowNet extracts information on how conditions have changed over the previous hour leading up to the initialisation time i.e. identifying whether convection is starting to develop or decay. The number of predictor channels is purposefully kept small, to reduce computational expense and training time.

SII-NowNet uses the input matrix to produce a 26 x 26 gridded nowcast over the same domain (see next section for dimension justification) i.e. it spatially coarsens the input data by a factor of 16 (Figure 5.2). The grid size of the nowcast therefore depends on the resolution of the geostationary satellite data. For example, using Himawari-8/9 BT data with a grid spacing of 2 km results in a SII-NowNet nowcast with a 32 km grid resolution, covering a region of 832 km x 832 km (94.6°E – 106.9°E, 6.4°S – 5.9°N).

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

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

SII-NowNet is trained separately for intensification and initiation, producing nowcasts with lead times at hourly increments for both. For each initialisation time, SII-NowNet produces a total

of 12 nowcasts, six for intensification and six for initiation (representing lead times of 1 to 6 hours).

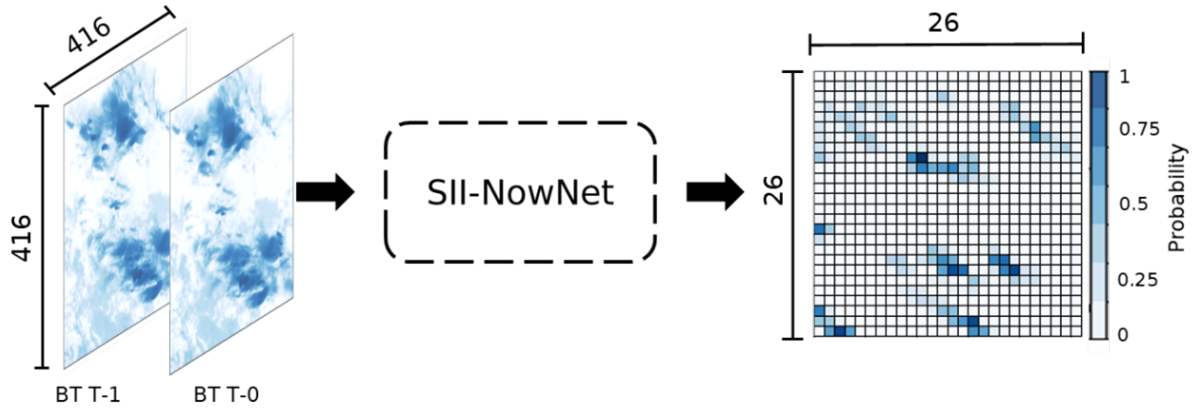


Figure 5.2: A diagram of the SII-NowNet nowcast process (applicable to both intensification and initiation predictions). The 416 x 416 x 2 input matrix (left) is input to SII-NowNet (middle), which outputs a 26 x 26 probabilistic nowcast over the same domain as the inputs (right).

5.2.3 Convective event observations

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

Both intensification and initiation events are identified on a 26 x 26 gridded domain (hence the dimensions of the SII-NowNet nowcast) using hourly data. Hourly data was subjectively chosen as it provided an appropriate timescale over which these events could be distinctly identified. Furthermore, the grid size was chosen to match the spatial scale of the initiation/intensification events, by empirically testing over multiple case study events.

Intensification identification

Figure 5.3 provides an example of intensifying convection events over Sumatra that are identified using the intensification method. For each individual grid (within the 26 x 26 gridded domain), convection intensification is identified by comparing two consecutive BT images, separated by

an hour (T-1 hours and T-0 hours, where T is the time of the intensification event), over the grid. Each BT pixel within the grid is defined as intensifying if it satisfies two criteria:

1. BT decreases by ≥20 K between T-1 and T-0
2. BT ≤235 K at T-0

As in chapter 4, in line with previous research, a threshold of 235 K is used in step 2 to try and isolate as much convection as possible (Crook et al. 2024; Feng et al. 2021; Machado and Laurent 2004; Roca et al. 2017). Finally, if the fraction of intensifying BT pixels within the grid is >=5%, the grid is labelled 1 for intensifying or 0 for not intensifying. This process is simultaneously repeated for each grid in the 26 x 26 gridded domain and then applied to all available BT image timesteps. This results in 17,326 intensification gridded samples for Sumatra, and 4,458 intensification gridded samples for Zambia.

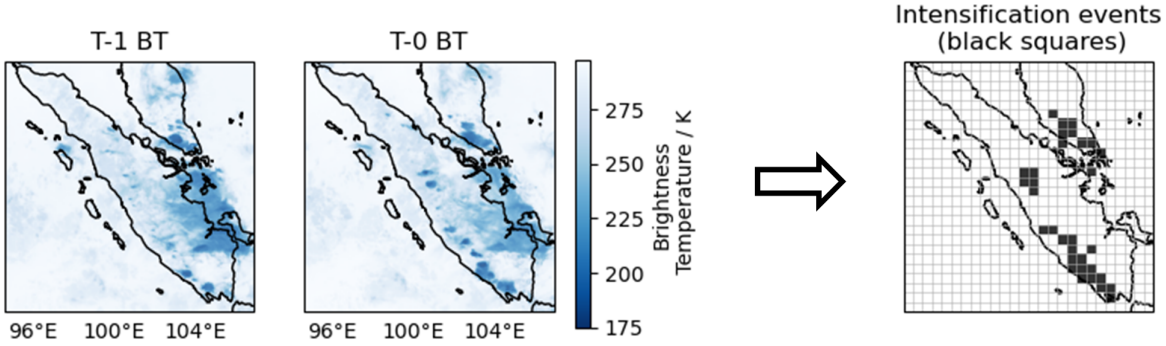


Figure 5.3: An example of intensifying convection along the south coast of Sumatra on 5th January 2019 at 18:00 LT. By comparing the T-1 BT and T-0 BT (left of arrow), the intensification identification method extracts only the intensifying convection, which it outputs on a 26 x 26 gridded domain (right of arrow).

Initiation identification

Convection initiation is also identified using two consecutive BT images, separated by an hour (T-1 hours and T-0 hours, where T is the time of the initiation event). To be classified as an initiation, the new convective cells have to be isolated from the convection that is already present (at T-1). The method follows two key steps, which are applied to each BT pixel within the nowcast domain (Figure 5.4 provides an accompanying example of the method):

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

of its size in the T-0 BT.

2. The estimated T-0 BT is subtracted from the observed T-0 BT to leave the remaining new convective cells that develop over the hour i.e. the convection.

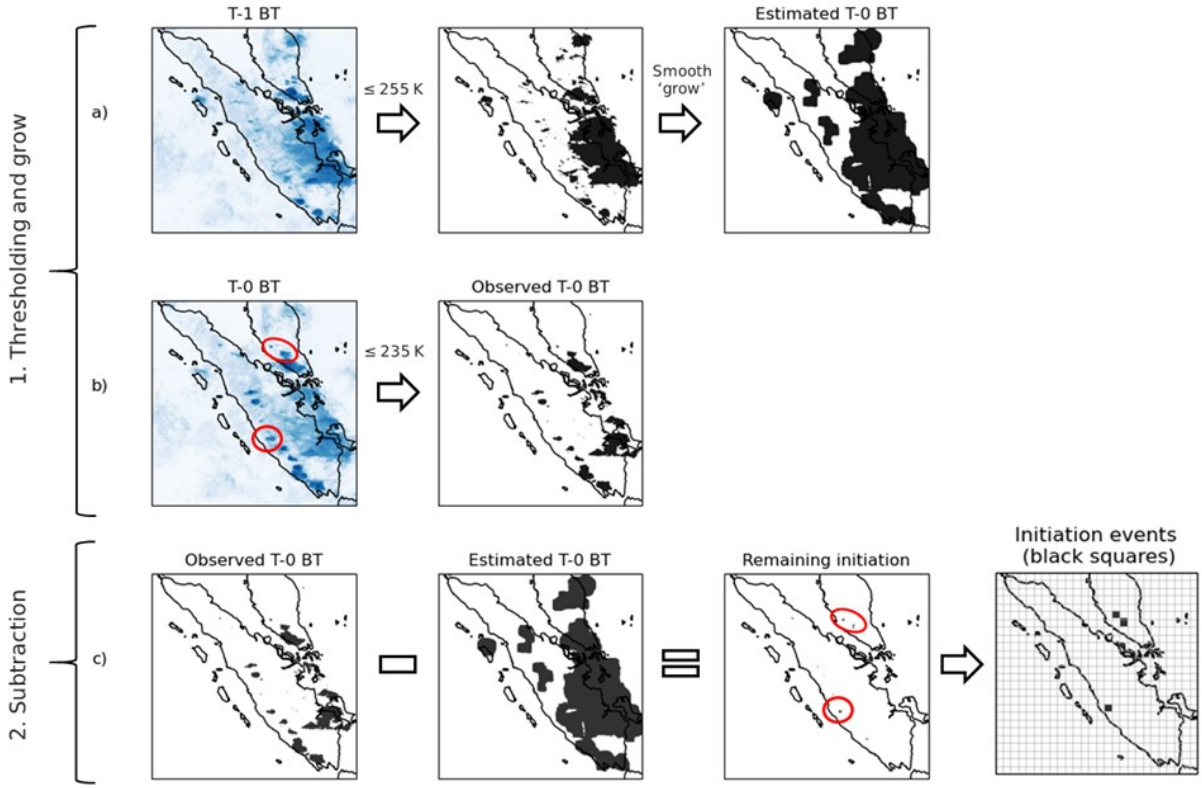


Figure 5.4: Using the same case as Figure 5.3, the two key processes for identifying convection initiation. a) The thresholding and then smoothing of the T-1 field to 'grow' the convection that is already present (T-0 BT estimate). b) The thresholding of the T-0 field to extract the convective regions only (T-0 BT observation). c) The final step showing the T-0 BT estimate field being subtracted from the T-0 observation field to leave only the isolated initiation (circled in red).

The convection growth simulation is performed by thresholding the T-1 BT at $\leq$255 K and then applying a smoothing filter (Figure 5.4a), which 'grows' the convection. A 255 K threshold is used to allow for regions where the convection may deepen and decrease in temperature by 20 K (the growth allowance value) over the hour and remain consistent with the intensification identification method. The T-0 BT observation is thresholded at 235 K to isolate only the convective regions (Figure 5.4b). The T-0 BT observation is then subtracted from the T-0 BT estimated field (Figure 5.4c).

This method assumes that the present convection in the T-1 BT will grow between T-1 and

T-0. However, if in reality the convection does not grow (e.g. dissipates), then subtracting the T-0 BT estimate from the T-0 BT observation will result in regions with negative values. To overcome this, all negative values are converted to 0, representing no convection initiation.

Finally, for each grid in the 26 x 26 gridded domain, if the fraction of initiation BT pixels is $\geq$5%, the grid is labelled as 1 for initiating or 0 for not initiating. This process is simultaneously repeated for each grid in the 26 x 26 gridded domain and then applied to all available BT image timesteps. This results in 17,326 initiation gridded samples for Sumatra, and 4,458 intensification gridded samples for Zambia.

This initiation identification method relies on applying a set of BT thresholds to isolate the newly developing convection. Figure 5.5 shows how the number of identified initiations decreases when a) the size of the smoothing kernel is increased and b) the growth allowance value is increased from 0 K - 30 K. Increasing the value of both parameters results in more isolated initiations being encompassed by the T – 0 BT estimate (Figure 5.4a), reducing the overall number of initiations. The rapid decrease in the number of initiations as the growth allowance increases (Figure 5.5b), shows that the method is highly sensitive, a limitation acknowledged by the author.

The difference in the number of identified initiations between Figure 5a and b highlights the large effect that the smoothing filter has on these events. This study applied a 20x20 pixel smoothing kernel to the T-1 BT image (Figure 5.4a). Both the kernel size and the growth allowance value were chosen after empirically testing a range of their values (across multiple case studies): a 20 K growth allowance and a 20x20 pixel kernel provided an acceptable balance between capturing cloud motion without encompassing too many nearby initiations.

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

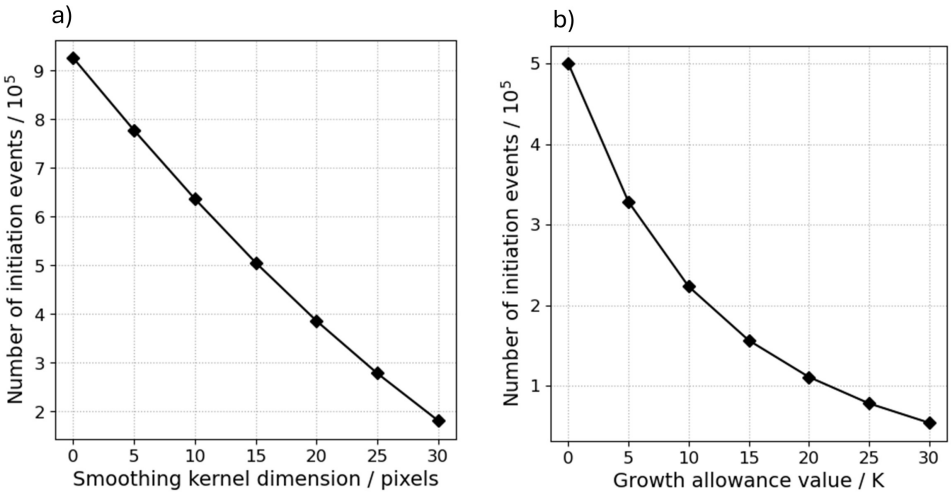


Figure 5.5: The number of events that are identified by the convection initiation identification methodology as the a) smoothing kernel dimension is varied and b) the growth allowance value is varied. In this study, a 20 pixel smoothing kernel and 20 K growth allowance were chosen.

Climatology of identified convective events

Table 5.1 shows the total number of observed events and non-events that are identified for initiation and intensification. Initiation events (1% of dataset) are observed less frequently than intensification events (14% of dataset). Both intensification and initiation have a clear diurnal cycle in their frequency, where event activity is minimal between 00:00 LT – 10:00 LT (Figure 5.6). Initiation events show a sharp peak in frequency at 14:00 LT that quickly declines into the afternoon, whereas intensification events are at maximum frequency approximately 2 hours later and continue to maintain significant signal during the afternoon and evening. Peak activity in the afternoon for both types of event is typical in many regions of the Tropics (Mori et al. 2004; Yang and Slingo 2001).

Dataset	**Events**	**Non-events**	**Balance**
Intensification	1,652,207	10,060,169	14%
Initiation	110.838	11,601,538	1%

Table 5.1: The total count of observed events/non-events for both convection intensification and initiation using hourly BT data over the study period (01/01/2018 – 31/12/2019)..

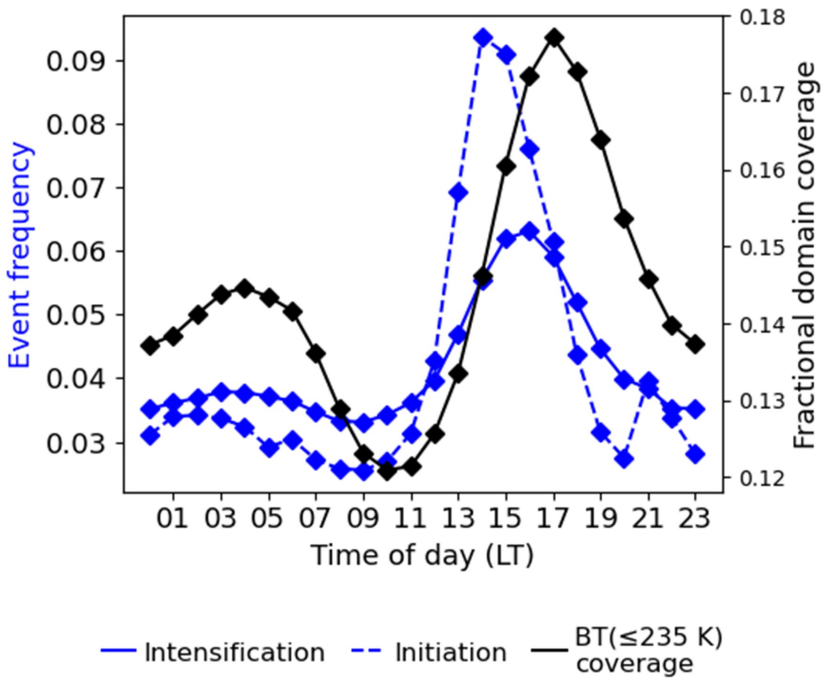


Figure 5.6: In blue, the frequency of observed events for convection intensification and convection initiation (Table 5.1), at each hour of the day (both curves are normalized and sum to 1). In black, the fraction of BT (≤ 235 K) coverage across the Sumatra domain at each hour of the day.

The peak coverage of BT (≤ 235 K) across the Sumatra domain arrives near 17:00 LT, approximately an hour after peak intensification and 3 hours after peak initiation. This aligns with the expectation that, as peak convection is reached in the late afternoon, maximum initiation/intensification is observed in the hours prior (when convection is building up).

5.2.4 SII-NowNet architecture

During the early development stages, SII-NowNet was based on a standard convolutional neural network. With this setup, one grid prediction was produced at a time, with only the grid's local BT data used as input. By iteratively producing grid predictions, this approach has the benefit of not being constrained to a domain size. However, the relatively simple architecture, combined with inputs that area restricted to local conditions, meant that the model produced a high number of false alarms for nowcasting initiation.

To increase the model's complexity, SII-NowNet was further developed using a U-Net architecture (Ronneberger et al. 2015), named due to the U-shaped path that the input takes through the model. The simple training routine and low computational demand of U-Net's makes them

a favorable choice for development in resource-constrained settings (e.g. regions in the Global South), as well as demonstrating nowcasting skill in previous studies (Lagerquist et al. 2021; Ayzel et al. 2020).

Figure 5.7 is a diagram of the U-Net design used for SII-NowNet. The 416 x 416 x 2 input matrix undergoes four encoding and decoding stages before processing into the desired output. The number of encoder/decoder stages was varied during the development stages of SII-NowNet's. Four stages was chosen as it provided enough complexity for the model to produce high skill scores, without causing it to overfit the data - a common problem in ML when too much complexity is built into a model.

During the first part of each encoding stage, the input (represented by the black arrows in Figure 5.7) is passed through two 2-D convolutional layers, each with a 3 x 3 filter, which aim to identify key features in the data. Attached to the end of each 2-D convolutional layer is a batch normalization function (to stabilize the network during training) and a ReLU activation function (Agarap 2019) (to enable non-linear patterns in the data to be learnt). In the second part of the encoder stage, the output from the two convolutions is passed through a max pooling function with a 2 x 2 filter. This downscales the output whilst retaining the most important features, resulting in a more compressed representation of the input data. However, downscaling can cause important information to be lost. To account for this, the output from the first part of the encoding stage (before max pooling) is saved and transferred back into the model at a later stage (during decoding). This is known as the skip connection and helps to retain fine spatial details. After 4 encoder stages the spatial dimensions of the original input data have been compressed to a 26 x 26 grid, and the dimension of the third channel is increased to 1024. This 26 x 26 x 1024 matrix is known as the latent space and represents the most important features that the model has identified from the input.

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

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

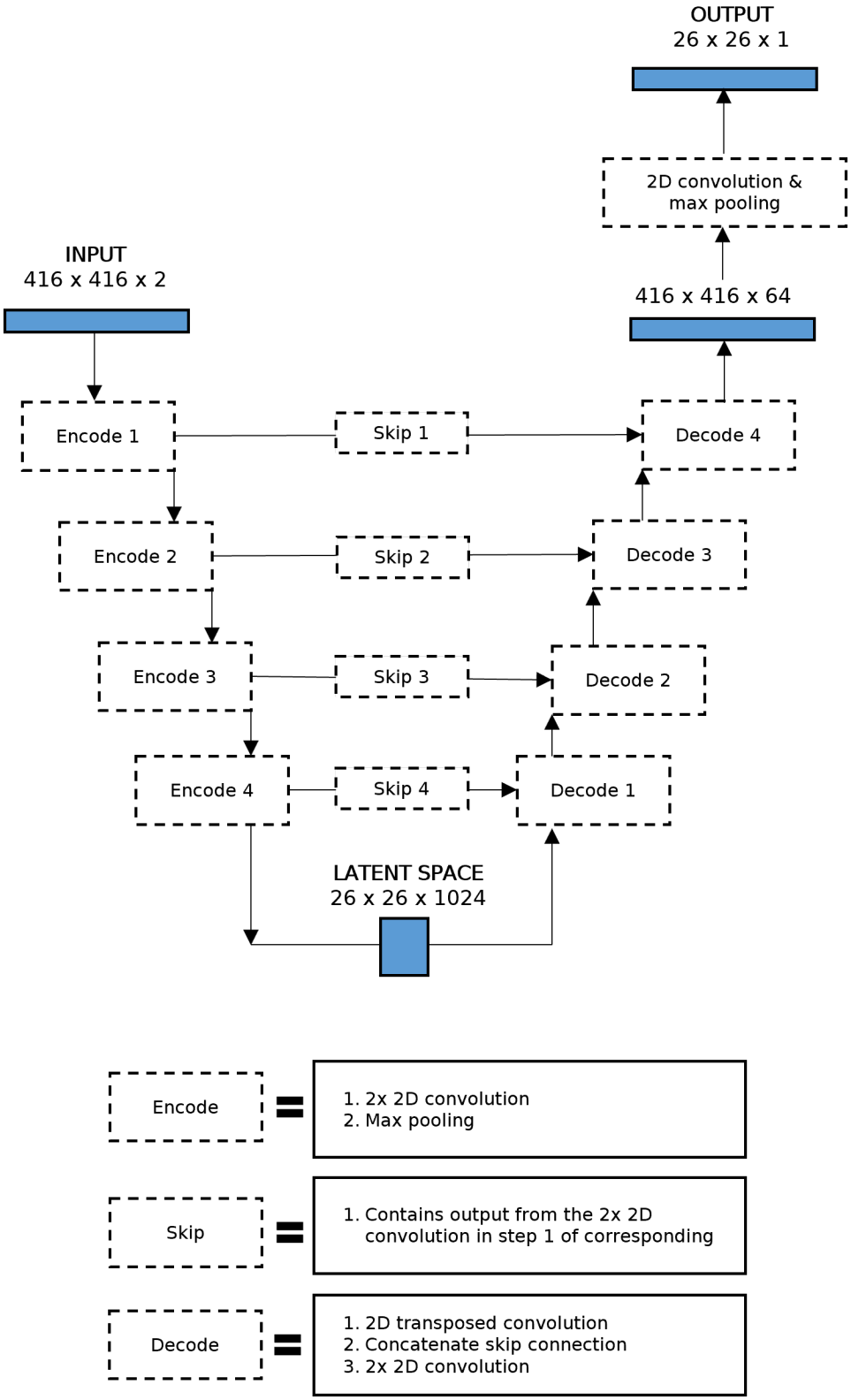


Figure 5.7: A diagram of the SII-NowNet model. A U-Net architecture with four encoder steps followed by four decoder steps, with skip connections in between. A final processing stage converts it to the correct dimension. The arrows represent the input matrix entering each step. Each blue square shows the shape of the input matrix as it passes through different stages of the network.

5.2.5 Training and testing data

SII-NowNet was trained across all times of day. For each hourly lead time (between 1 - 6 hours), for both intensification and initiation, a unique version of SII-NowNet was trained. This results in 12 unique SII-NowNet models presented in this study i.e. 6 future hours for the two predictands (initiation and intensification). Each model was trained for an upper limit of 10 epochs (i.e. 10 full passes through the entire training dataset) using a binary cross entropy loss function. This function measures the difference between the predicted probability and the actual label, and is typically used in binary classification problems. The Adam optimizer (learning rate = 0.001) was used during training to efficiently update SII-NowNet's parameters. However, the maximum number of required training epochs across all lead times for intensification and initiation was 8. To ensure no overfitting to the training data, an early stopping routine was applied to the validation loss function (a 20% random subset of the training data was used as the validation data). If the loss value stayed above the lowest previous loss value for more than 1 epoch, training ended and the model returned to its state at the previous lowest loss value.

To account for the imbalance in SII-NowNet's training and test data (Table 5.1), the training set was re-balanced to contain 40% events (for both initiation and intensification) and tested on data with the original climatological balance. However, when the proportion of initiations was increased in the training set, SII-NowNet became overconfident and produced a higher number of false alarms during testing. A class weighting was also tested, which gave greater 'reward' to SII-NowNet when it correctly predicted a rare event (e.g. initiations). However, overall, the best results were achieved when maintaining the climatological balance for both the training and testing datasets, and without class weighting.

For developing SII-NowNet over Sumatra, the available 17,326 initiation and intensification samples were divided into a training set and a test set. 8,661 training samples were taken from 2018 and used during the training process for model learning, whilst 8,665 test samples were taken from 2019 and were used for model verification only. Therefore, the test samples were unseen by the model and excluded from training. For developing SII-NowNet over Zambia, 1,676 training samples were taken from February 2024 for training, whilst 2,782 samples were taken from March 2024 for testing.

5.2.6 Permutation importance

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

5.2.7 Baseline

STEPS, climatology and persistence (see section 3.3 for persistence description) are used to assess whether SII-NowNet has any useful skill.

Using the same methods as in chapter 4, this study produces 20-member ensemble STEPS nowcasts of BT for 1–6 hours of lead time. The initiation and intensification identification methods (section 5.2.3) are then applied to the STEPS BT nowcasts (i.e. simply replacing the BT observations with the STEPS nowcasts at the same valid time), resulting in corresponding STEPS nowcasts for initiation and intensification.

In this study SII-NowNet's intensification nowcasts are evaluated against intensification persistence. Intensification persistence is defined as the T=0 intensification observation (on the same 26x26 grid), which is then used as the prediction for future lead times.

Given the large amplitude of the diurnal cycle of convection in the Tropics (Yang and Slingo 2001), this study defines climatology as the expected average number of convective events for a given time of day. Climatology is calculated separately for intensification and initiation at each hour of the day by computing the fraction of observed events from the sum of observed events and non-events (i.e. calculating the climatological mean at each hour).

5.3 Results

5.3.1 Qualitative analysis of SII-NowNet

Figure 5.8 presents a case study for SII-NowNet intensification and initiation nowcasts, valid at 20:00 LT on 7th March 2019 over Sumatra, Indonesia. This case was chosen because it highlights both the strengths and weaknesses of SII-NowNet. Contouring has been applied to all SII-NowNet nowcasts to provide a smoother probabilistic field (compared to the raw gridded output in Figure 5.2). Different colour bar scales have been used for intensification and initiation nowcasts to account for their differing probability ranges at each lead time.

At 1 hour lead time, SII-NowNet's intensification (Figure 5.8b) and initiation (Figure 5.8i) nowcasts produce regions of high probability that align with the observed intensification (Figure 5.8a) and initiation (Figure 5.8h) events e.g. northern Sumatra for both nowcasts. By producing these distinct regions of high probability, it highlights SII-NowNet's ability (at this lead time) to clearly discriminate between events and non-events. However, the initiation nowcast also provides an example of missing an initiation event (single grid, northeast Sumatra coastline), highlighting the challenge in capturing isolated events.

At 2 hours' lead time, SII-NowNet's intensification and initiation nowcasts continue to capture the observed events, but with a drop in confidence. The intensification nowcast (Figure 5.8c) still produces high probabilities that align with the bulk of observed events (e.g. south and southwest of Sumatra, Figure 5.8a), however, it now produces a less-than-30% probability near

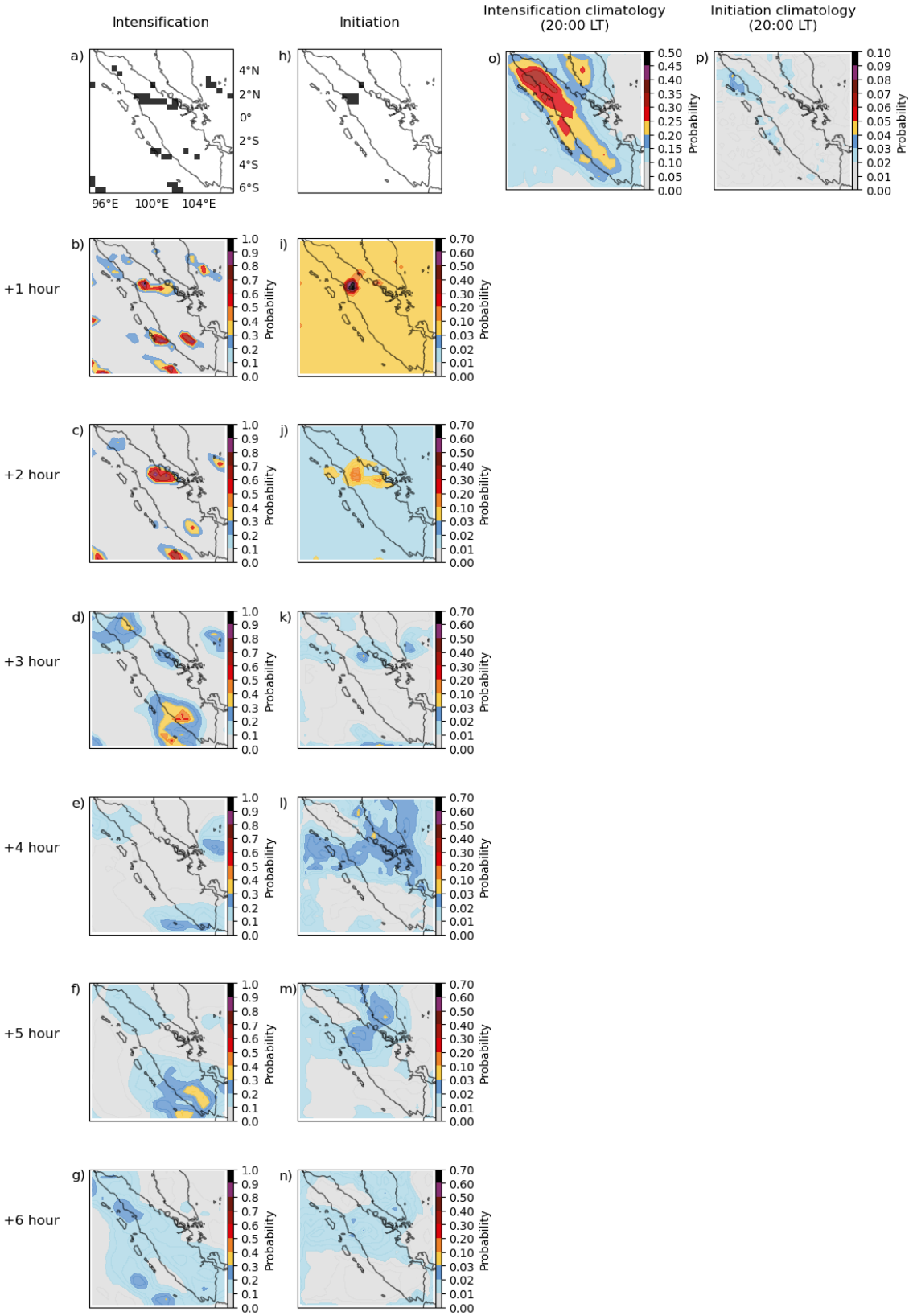


Figure 5.8: A case study of a) intensification and h) initiation events over Sumatra at 20:00 LT on 7th March 2019. b-g) and i-n) show the probabilistic SII-NowNet nowcasts for intensification and initiation, respectively, for 1 – 6 hours' lead time. Each nowcast is valid at 20:00 LT. o) and p) show the climatological means for intensification and initiation, respectively, at 20:00 LT. Different colour bar scales have been used for intensification, initiation and climatology to account for their differing probability ranges.

the southwest coast of Sumatra, where events are observed. Similarly, the initiation nowcast (Figure 5.8j) continues to produce the highest probabilities over northern Sumatra (similar to Figure 5.8i), which align with the observations; however, the maximum probability has dropped to 20%. Furthermore, the initiation nowcast contains a small region of higher initiation probabilities to the east, which does not align with the observations (i.e. it produces a false alarm here).

Beyond 2 hours' lead time, SII-NowNet's nowcasts increase in blurriness and decrease in maximum probability. In particular, the initiation nowcasts (Figure 5.8k-n) show a rapid drop in probability over this period, along with a reduction in precision. The intensification nowcast at 3 hours' lead time (Figure 5.8d-e), although blurry, still appears to show some precision, aligning with the observed intensification over the south, northeast and northwest of Sumatra. However, for 4 – 6 hours' lead time, the regions of higher intensification probability become much smoother, making it difficult to gauge specifically where convection will intensify.

For initiation nowcasts between 1 – 3 hours' lead time (Figures 5.8i-k), the minimum background probability reduces from less-than-10% to less-than-1% (before staying consistent beyond 3 hours). SII-NowNet is becoming more uncertain about its predictions of the (far less frequent) initiation events as lead time increases. To compensate for this uncertainty SII-NowNet reduces its probability values.

Figures 5.8o-p show the climatological mean for intensification and initiation at 20:00 LT. The highest intensification likelihoods are observed over the northwestern coastline of Sumatra - capturing the observed intensifications over this region, but also producing a number of false alarms. This pattern can be explained by convection that develops over the afternoon and continues to deepen into the evening, with propagation offshore expected into the evening (Mori et al. 2004). The initiation climatology at this hour produces widespread low likelihoods across the domain, missing the initiation over the northwestern coastline of Sumatra. At this time of day, far fewer initiations are occurring (Figure 5.6) because, on average, convection has already developed and is reaching maturity at this time (Mori et al. 2004).

5.3.2 Quantitative analysis of SII-NowNet

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

Figure 5.9 shows the ROC curves (a and b) and reliability diagrams (c and d) for SII-NowNet, STEPS and climatology nowcasts for 1, 2 and 3 hour lead times only (to avoid too many lines). The 4 hour lead time intensification result has also been added to the reliability diagram (d) to show its limit of capability. To account for SII-NowNet producing initiation nowcasts with varying probability ranges at different lead times (e.g. Figure 5.8i-n), their ROC curves are calculated across different probability ranges. Probability ranges of [0, 1], [0, 0.5], [0, 0.1], [0, 0.1], [0, 0.1], [0, 0.1] were used for 1 – 6 hour lead time initiation ROC curves. For intensification nowcasts, all ROC curves were calculated using a probability range of [0, 1].

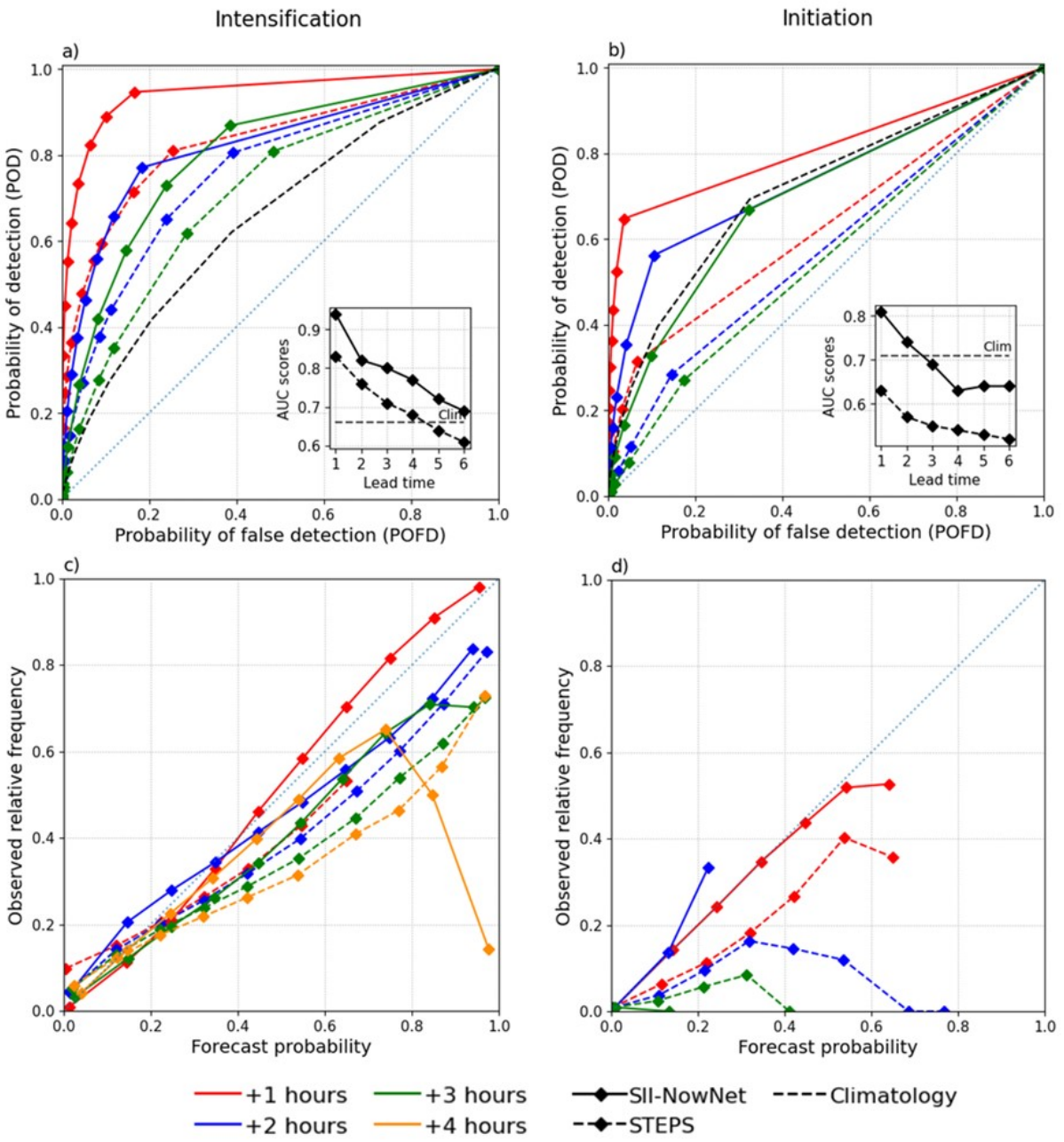


Figure 5.9: ROC curves for SII-NowNet, STEPS and climatology a) intensification and b) initiation nowcasts, containing curves for each lead time. The inset plots contain the AUC scores at each lead time, with the climatology shown by the horizontal dashed line. Reliability diagrams for SII-NowNet and STEPS c) intensification and d) initiation nowcasts, containing curves for each lead time. By definition the climatology reliability line lies on the x=y line. All analyses are generated using 8,665 unseen samples from 01/01/2019 – 31/12/2019. All plots contain a dashed x=y line for reference.

For intensification, SII-NowNet's ROC curves shift closer to the dashed x=y line (no skill) as lead time increases, resulting in decreasing AUC scores (inset panel Figure 5.9a) with lead time (0.94 to 0.69). At each lead time shown, SII-NowNet's ROC curves produce a greater AUC score than STEPS, reflecting a better nowcasting performance. At 6 hours' lead time, SII-NowNet's

AUC score is only 4% greater than climatology. In terms of reliability, SII-NowNet produces more reliable intensification nowcasts (i.e. closer to x=y) than STEPS for 1, 2 and 3 hours' lead times. At 4 hours' lead time SII-NowNet drastically over-predicts events at probabilities greater than 0.75, generating false alarms. 2% of SII-NowNet's intensification predictions are $\geq$0.5 at 4 hours' lead time (reducing to 0.1% at 6 hours), compared to 10% at the 1 hour lead time. Although the AUC scores show that SII-NowNet's intensification nowcasts outperform both STEPS and climatology over 1 – 6 hours' lead time, the poorer reliability beyond 3 hours' lead time indicates this as a reasonable limit of capability.

For initiation, the AUC scores (inset panel Figure 5.9b) produced by the ROC curves are less than the intensification scores for each corresponding lead time i.e. SII-NowNet is less skilful at predicting initiation events. The initiation scores decrease from 0.82 to 0.64 between 1 – 4 hours' lead time. For 4 – 6 hours, SII-NowNet's AUC scores remain approximately constant. SII-NowNet's initiation nowcasts outperform STEPS at each lead time but drops below the climatology between 2 – 3 hours' lead time. This represents a 2 hour lead time limit of capability for SII-NowNet's initiation nowcasts. This limit is also reflected in the reliability curves (Figure 5.9d), which show that SII-NowNet produces more reliable initiation nowcasts for 1 – 2 hours lead time than STEPS. For 1 hour lead time nowcasts these reliable predictions lie between a probability window of 0 – 0.6. For 2 hour lead time nowcasts the reliability probability window is approximately 0 – 0.15. For predicted probabilities between 0.15 – 0.22, SII-NowNet starts to under-predict initiation events at this lead time (the reliability curve tends towards the y-axis). At 3 hours' lead time, SII-NowNet's initiation nowcasts produce reliability curves that only extend along the x-axis, meaning that it is predominantly producing false alarms.

In general, SII-NowNet produces lower probabilistic values when nowcasting initiation compared to intensification (as exemplified in Figure 5.8). This suggests that the model is compensating for the relatively lower frequency of initiation events by associating its predictions with more uncertainty i.e. lower likelihood.

SII-NowNet's intensification nowcasts can also be compared against a persistence nowcast, which, on nowcasting timescales, can often have reasonable skill (as shown in chapter 4). Due to the nature of initiations, it is not possible to produce an initiation persistence nowcast (initiations are one-off events that cannot persist), so only a persistence of intensification is presented (Figure A.1 shows persistence for the same case study as in Figure 5.8). At each lead time

the SII-NowNet intensification nowcasts have been thresholded using an optimum probability threshold value (0.2, 0.1, 0.2, 0.2, 0.2 and 0.2 for 1 – 6 hours' lead time), to produce a set of deterministic nowcasts. The optimum threshold for each lead time is determined by using the threshold value from the corresponding ROC curve that produces the maximum POD – POFD difference. Figure 5.10 shows the accuracy, POD and 1–FAR of the deterministic SII-NowNet and persistence nowcasts of convection intensification (1–FAR has been used so that all skill scores of 1 represent a perfect score).

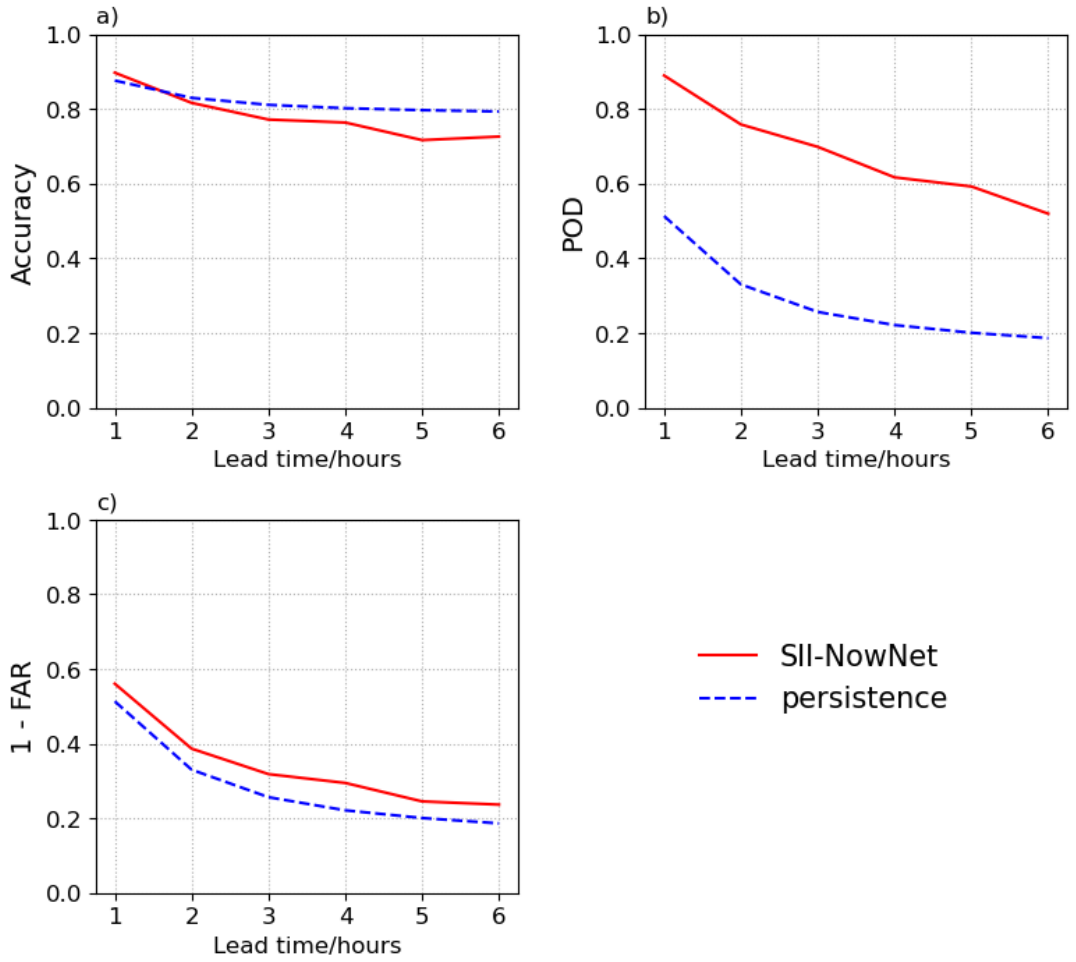


Figure 5.10: Mean scores for a) accuracy, b) POD and c) 1-FAR across each lead time for intensification nowcasts produced by SII-NowNet and persistence using the same test data as in Figure 9.

In terms of accuracy (Figure 5.10a), persistence and SII-NowNet perform similarly up to 4 hours, beyond which persistence has a greater accuracy score by approximately 0.1. The persistence nowcast gains the majority of its accuracy from correct negatives, whereas SII-NowNet's

accuracy is derived from a combination of correct negatives and hits. This is shown in the POD scores (Figure 5.10b), where SII-NowNet outperforms persistence across all lead times by 0.45. Although persistence will often be correct in predicting no event, when an event does happen it is 45% less likely to produce a hit. On the other hand, SII-NowNet has learned the conditions within the predictor fields that are associated with an intensification event, meaning it will more often produce a hit. Even though SII-NowNet produces a greater number of intensification events, it still maintains a greater 1 – FAR score than persistence by approximately 0.05 across all lead times. This means that, when both nowcasts predict an intensification event, persistence is 5% more likely to produce a false alarm than SII-NowNet.

5.3.3 Interpreting SII-NowNet

To gain understanding into SII-NowNet's decision-making, it is possible to analyse feature maps that are produced by the model as input passes through each stage of its architecture. A feature map is the output of a convolutional filter within one of the encoder/decoder steps (Figure 5.7), which shows where the filter is most active across the input i.e. highlighting the spatial patterns that are being identified at that layer. This is measured by relative importance - the higher the relative importance, the more attention the filter is giving to that location.

Figure 5.11 provides the feature heatmaps of relative importance produced by SII-NowNet at five different stages through its U-Net architecture, using BT input data from 27th December 2019 (Figure 5.11a-b). Each heatmap is overlayed ontop of the T=0 BT input image (Figure 5.11b) and the final SII-NowNet nowcast's (for a 1 hour lead time) are shown for initiation (Figure 5.11h) and intensification (Figure 5.11n). For initiation, the encoder 2 stage (the beginning of the U-Net) appears to give greater importance to fine-scale, colder BT features that are present at T=0 (Figure 5.11c), and less focus on the larger regions of cold BT. Feature importance is more blurred at encoder 4, where the model appears to being identifying non-convective regions across the domain (Figure 5.11d). In the decoder 2 stage, SII-NowNet has identified distinct, small-scale features of high importance (Figure 5.11e) and begins to show resemblence to the final prediction (Figure 5.11h). SII-NowNet then appears to identify the regions of convection surrounding these small-scale features in the decoder 4 stage (Figure 5.11f). Finally, it appears to use the surrounding convective features as a mask to filter out some of the decoder 2 features in the final convolution stage (Figure 5.11g). The output of the final convolution can then be easily mapped to the SII-NowNet's initiation prediction (Figure 5.11h).

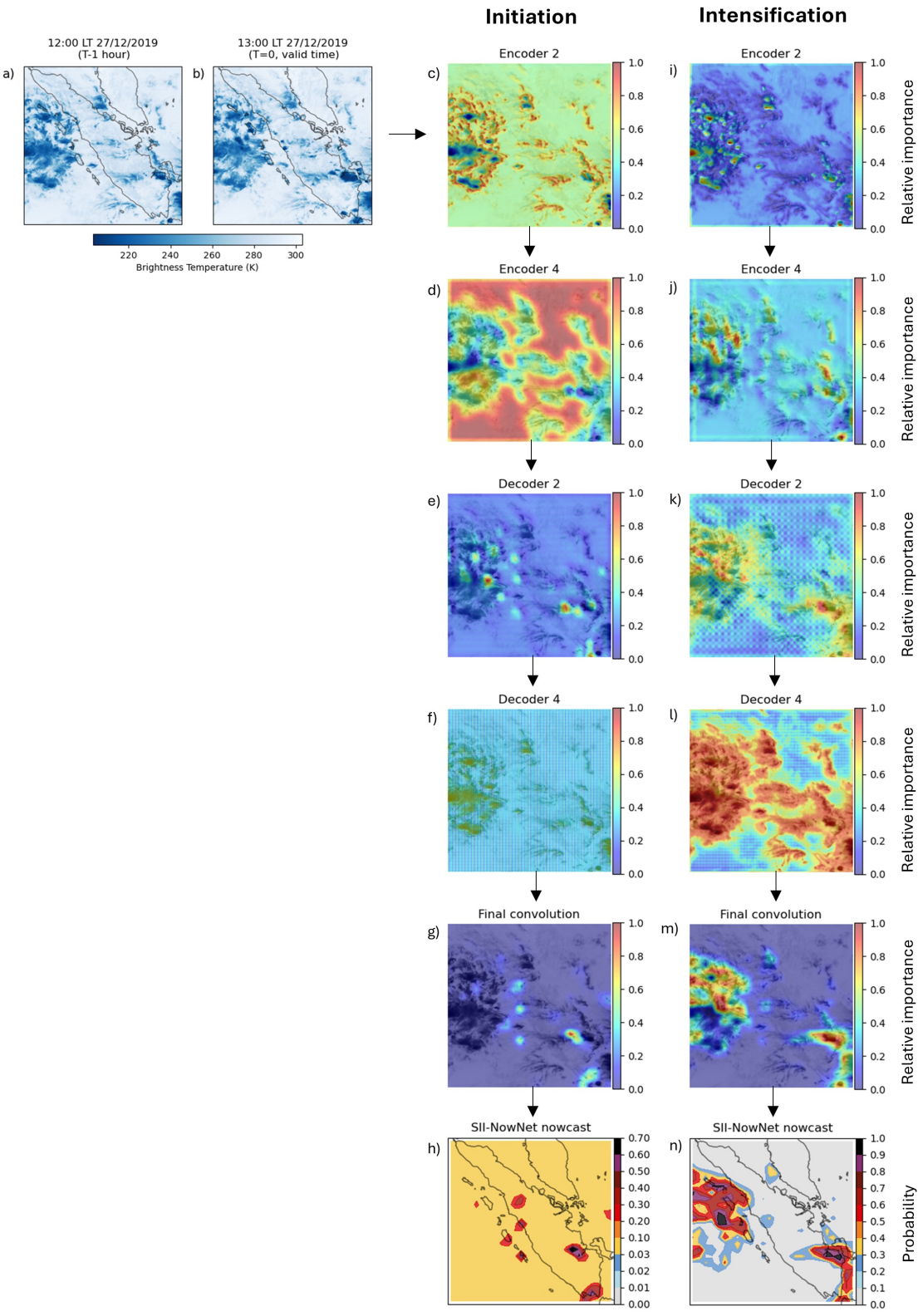


Figure 5.11: Heatmaps of relative importance at intermediate stages of SII-NowNet's U-Net architecture (Figure 5.7), using a case study from 27th December 2019. a) and b) show the T-1 and T-0 BT satellite images, respectively, that are used as input to SII-NowNet for this case study. c) - g) and i) - m) show the heatmaps of relative importance for initiation and intensification, respectively, produced by SII-NowNet as the input data passes through different stages of the U-Net architecture. h) and n) show SII-NowNet's final 1 hour lead time nowcast for initiation and intensification, respectively.

For intensification, the encoder 2 stage also appears to identify small-scale features with the coldest BTs (Figure 5.11i), within the larger envelope of cold BT cover (Figure 5.11b). Over the encoder 4 and decoder 2 stages, SII-NowNet continues to give importance to these cold BT regions, but with more focus on regions of deepening BT (Figure 5.11j-k) e.g. over the southeast of Sumatra and the northwest of the domain (Figure 5.11b). At the decoder 4 stage, SII-NowNet's feature map produces a large, blurred region of relative importance, appearing to encompass all cold cloud top temperatures. Similar to the initiation process, SII-NowNet appears to then use the feature map from decoder 2 and decoder 4 to filter out non-intensifying regions in the final convolution (Figure 5.11m), and produce a feature map that can be easily mapped to the intensification nowcast (Figure 5.11n).

To test SII-NowNet's ability in capturing the location of convective features, its predictions can be compared for different orientations of the input data. An example of this can be seen in Figure 5.12, which shows SII-NowNet's 1 hour lead time nowcasts from the same case study as Figure 5.11, but with the input data rotated by 0°, 90°, 180° and 270°. Figure 5.12a-d shows the standard SII-NowNet nowcasts without a rotated input, as a reference baseline. For the 90° rotated input (Figure 5.12e-f), SII-NowNet's initiation nowcast (Figure 5.12g) produces the majority of its high likelihoods in the southwest of the domain. Its region of highest likelihood aligns with the same BT feature that is identified at 0°, suggesting that SII-NowNet has been able to apply the same identification methods in a new location. This is also consistent throughout the 180° (Figure 5.12k) and 270° (Figure 5.12o) rotations, where, for the highest initiation likelihoods, SII-NowNet consistently identifies the same features as they are moved across the domain. For the intensification rotations (Figures 5.12d,h,l,p), SII-NowNet produces an almost identical distribution of likelihoods rotated in each new orientation, again highlighting its capability to capture the convection as it moves to new locations across the domain.

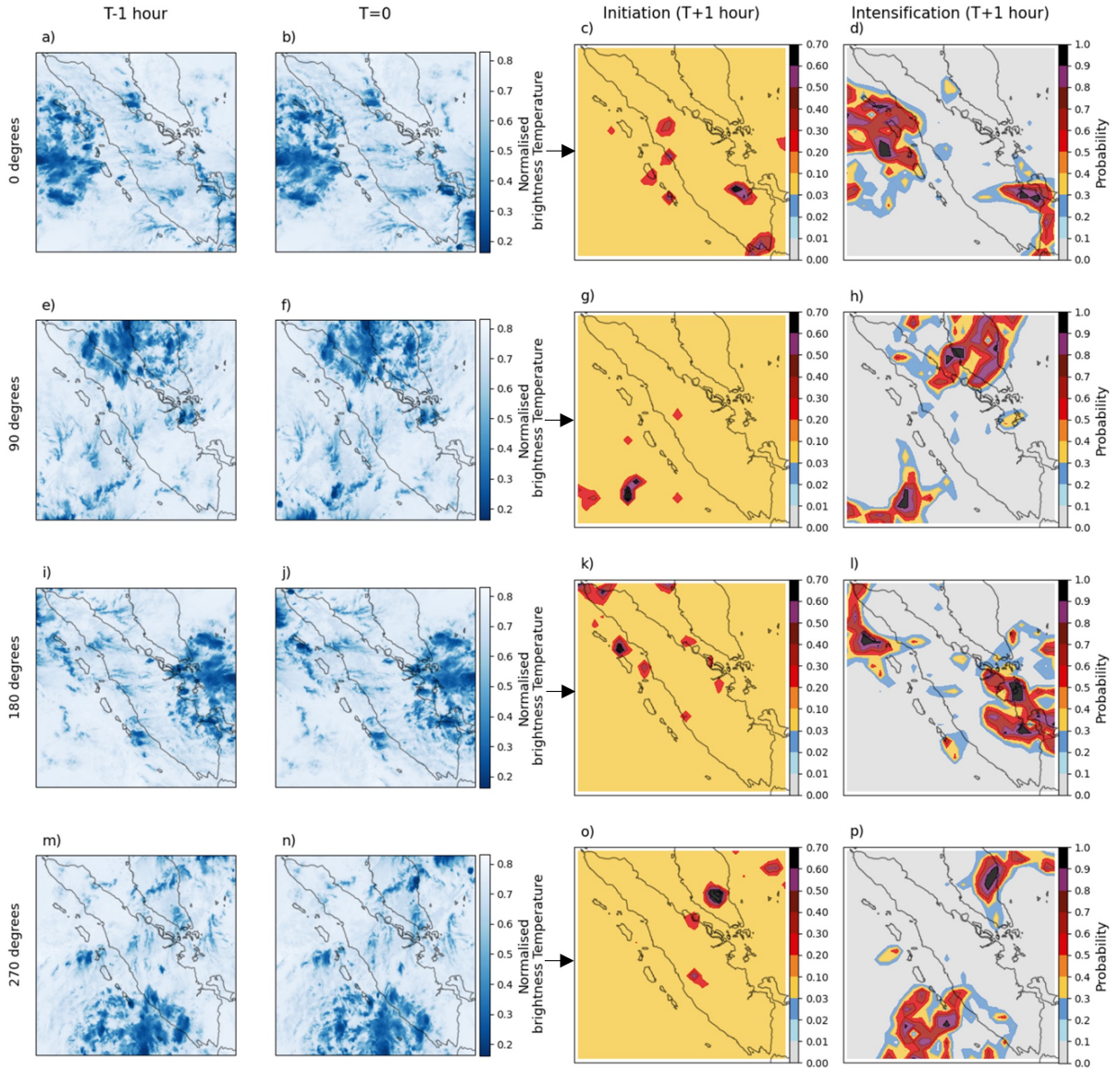


Figure 5.12: Each row shows the T-1 and T=0 BT input images, along with the corresponding SII-NowNet nowcasts (at a 1 hour lead time). The BT input data has been rotated for 0°, 90°, 180° and 270° rotation for each row, using the same case study as Figure 5.11.

By rotating the orientation of the 0° prediction (Figures 5.12c-d) (the control runs) to match the orientation of each rotated prediction (Figures 5.12 g,k,o), and then calculating the difference, any tendency that SII-NowNet has to focus its higher likelihoods over particular regions can be highlighted. Repeating this for multiple examples (as in Figure 5.12) provides a way to detect anomalously large differences across the domain. Areas of positive anomalies highlight SII-NowNet's tendency to produce higher likelihoods of initiation/intensification over that region. Figure 5.13 shows a map of mean positive anomalies of the difference between the control run and the rotated nowcasts, calculated across 2,000 SII-NowNet nowcasts (at a 1 hour lead time)

randomly selected from 2019. The positive anomalies are calculated using the mean z-score at each SII-NowNet grid point. The z-score is given by:

$$z = \frac{x - \mu}{\sigma}, \tag{5.1}$$

where x represents the value of a given grid, μ represents the mean value at the grid point and σ represents the standard deviation at the grid point. Larger values of z represent greater anomalies.

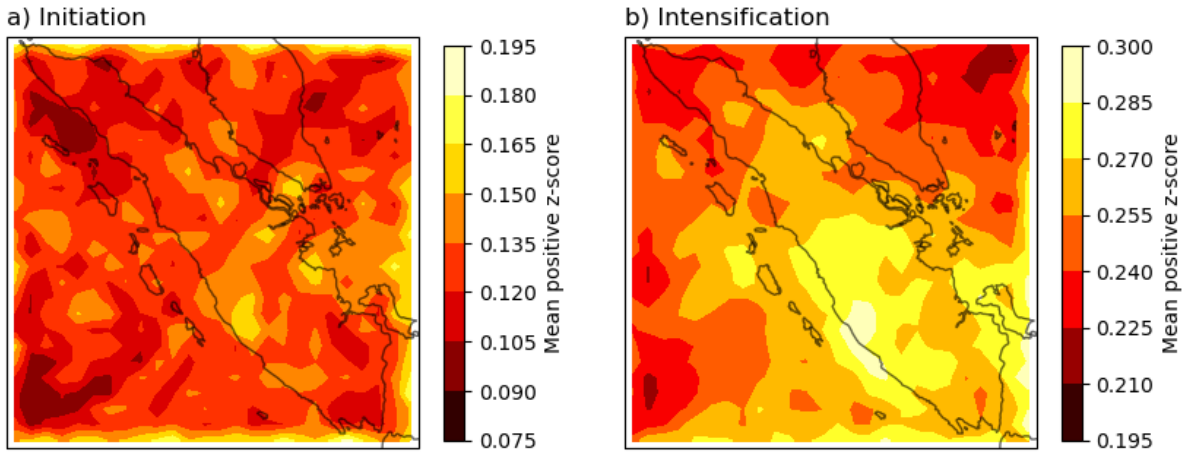


Figure 5.13: Mean positive z-scores for the difference between SII-NowNet's control runs (0° rotated input) and the corresponding rotated inputs for 90°, 180° and 270°, across 2,000 cases in 2019, for a) initiation and b) intensification. High values represent higher anomalies, which show SII-NowNet's tendency to predict higher likelihood of an initiation/intensification event.

Figure 5.13a shows the positive anomalies for the initiation nowcasts. The lack of clear widespread signal suggests that SII-NowNet's predictions do not have a strong tendency to produce high likelihoods over particular regions. However, the areas with the highest anomalies can be seen over the island of Sumatra and near the south coast of Malaysia. In particular, the largest anomalous region can be seen over the west coast of Sumatra, which aligns with some of the island's orography. High anomalies are also seen around the edges of the domain for the initiation nowcasts. This is likely at artifact caused by the identification process - propagating convective systems entering the domain may be incorrectly identified as new initiations. This highlights a weakness of the identification method, which could be solved by first applying it to a wider domain and then cropping the identified events down to the smaller study domain.

For intensification (Figure 5.13b), the positive anomalies highlight a clear region of tendency of

SII-NowNet to produce higher likelihood predictions over southern Sumatra. In particular, the highest anomalies are concentrated over the west coast of Sumatra (aligning with those in Figure 5.13a) where the island's orography is located. Over the sea, the intensification signal contains less noise than for initiation, suggesting that SII-NowNet has a clearer tendency to not produce higher likelihoods there. This pattern of high anomalies over the land and low anomalies over the sea could be explained by SII-NowNet learning about the underlying geographical controls that link to enhanced convective activity.

5.3.4 Sensitivity testing the number of training samples

Many operational meteorological services across the Tropics have limited access to large data storages or GPU compute, meaning they are unable to run data-intensive models. It is therefore important to test the minimum data requirement of ML nowcasting models for operational practicality. To understand the lower limit of training samples that SII-NowNet requires to maintain acceptable performance, the 1 hour lead time version was re-trained 6 times (for both initiation and intensification), each time halving the number of training samples that were used. Figure 5.14 shows the AUC scores and fraction of original training time (58 minutes and 3 seconds) for each halving iteration. For intensification a 94% reduction in the number of training samples, from a year of data to approximately three weeks of data (8,661 to 541), results in an 11% reduction in SII-NowNet's AUC score and a 99% reduction in training time. However, further reduction of training samples from 541 to 270 results in a rapid drop in AUC score from 0.84 to 0.5. This provides an approximate limit of training samples for achieving acceptable SII-NowNet intensification performance with two orders of magnitude greater speed. For initiation, SII-NowNet reduces more steadily as the number of samples reduces. A 75% reduction in the number of training samples, from a year of data to approximately three months of data (8,661 to 2,165), results in an 11% reduction in SII-NowNet's AUC score and a 75% reduction in training time. At this point SII-NowNet's AUC score is approximately equal to the initiation climatology (Figure 5.14b), providing an approximate limit of training samples for initiation.

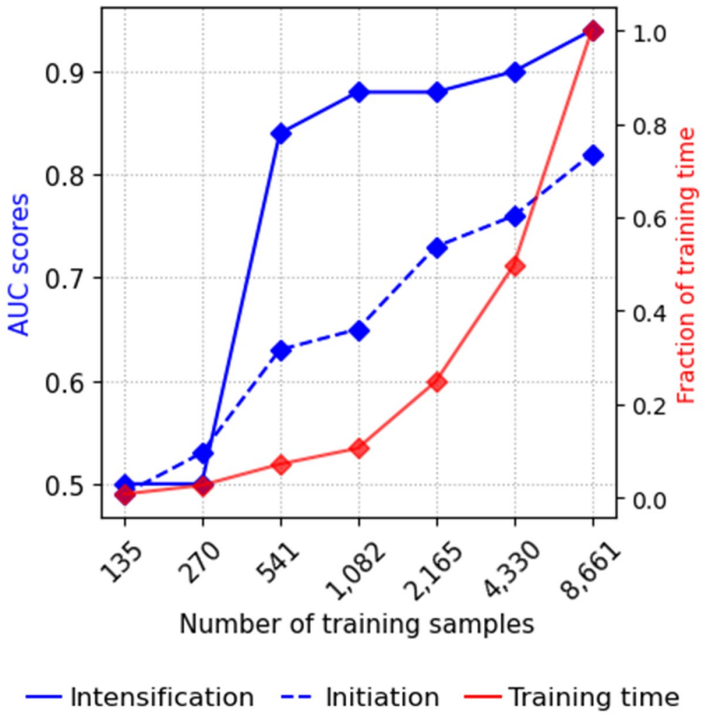


Figure 5.14: *Left axis* shows the AUC skill of 1 hour lead time SII-NowNet initiation (blue dashed) and intensification (blue solid) nowcasts as the number of training samples are continuously halved from the 8,661 (a year of data) to 135 (approximately three weeks). *Right axis* shows the fraction of time it took to train the model with this number of training samples, compared to the full training dataset (58 minutes and 3 seconds).

5.3.5 Input permutation testing

To gain insight into SII-NowNet's decision-making process, a permutation test is used to provide information on the relative influence that each of its inputs has on its predictions. Figure 5.15 shows the relative reduction in AUC score for SII-NowNet (for lead times that are within the limit of usefulness only) after each input predictor (BT T-0 and BT T-1) is, in turn, replaced with a Gaussian noise field. A greater relative reduction in AUC is associated with the predictor having a greater influence on the prediction (i.e. greater importance). In general, as lead time increases (and SII-NowNet skill decreases), there is a reduction in the influence of both predictors on SII-NowNet's nowcasts of initiation and intensification. As lead time increases, the BT data becomes less relevant for convective activity at the nowcast valid time, hence the predictors' influence on SII-NowNet reduces. The BT T-0 (the latest BT image) input produces the greatest reduction in AUC for SII-NowNet across all lead times. This is to be expected as it contains the most up-to-date information on cloud activity over the region. However, across all lead times, the BT T-1 permutation produces a minimum of 20% reduction in AUC score.

This significant reduction suggests that, although not as important as the latest BT image, information in BT data from the previous hour is still useful for predicting the likelihood of convection development for the next few hours.

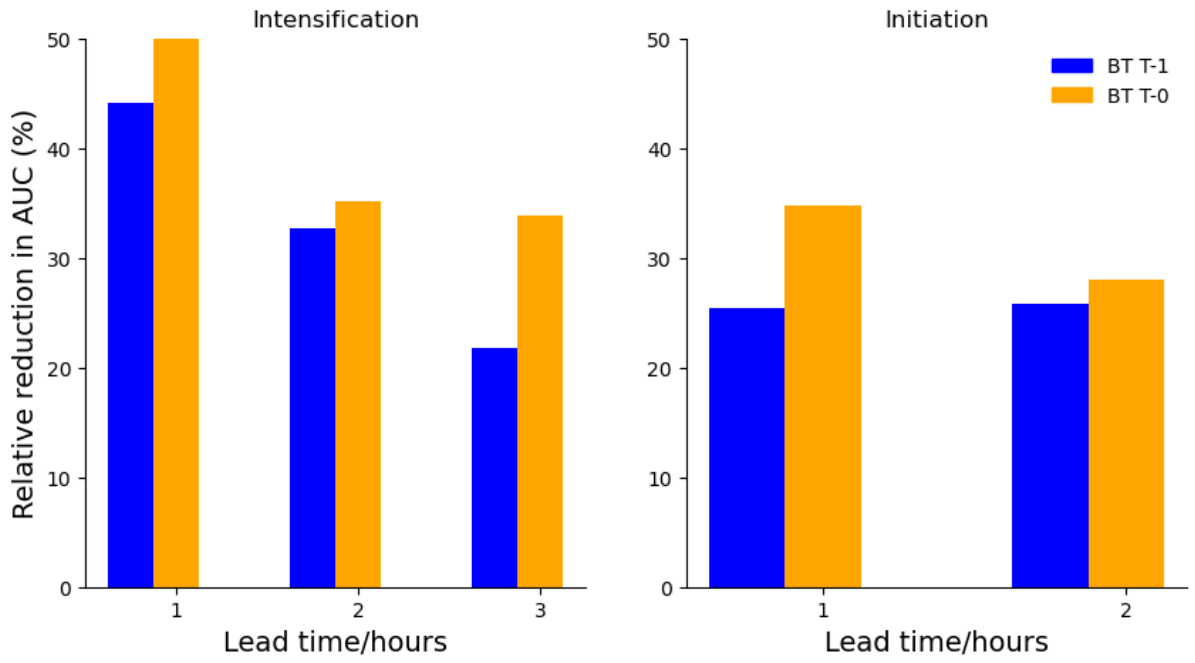


Figure 5.15: The mean relative reduction in the AUC score of SII-NowNet when each variable is, in turn, replaced with a Gaussian noise field. The Gaussian noise replacement was repeated 20 times for each variable across each lead time that is considered skilful.

5.3.6 SII-NowNet applied to Zambia

To provide evidence of SII-NowNet's applicability to other Tropical regions, its performance is evaluated over Zambia (without any re-training or region-specific tuning). Furthermore, the performance of SII-NowNet-Zam – a new version of SII-NowNet that has been given additional training on 1 month (1,675 samples) of BT data over Zambia from February 2024 – is presented for comparison. Therefore, SII-NowNet-Zam is trained on 84% BT data from the Sumatra domain and 16% for the Zambia domain.

Zambia experiences stronger seasonal patterns of rainfall compared to Sumatra. During the wet season (November - April), the ITCZ is approximately over Zambia, providing favourable conditions for convection (Rouault et al. 2024). In a typical day during the wet season, convection builds throughout the afternoon with peak rainfall observed in the late afternoon/early evening (Jury 2024). This pattern is reflected in Figure 5.16, which shows the diurnal cycle

of intensification and initiation events over the Zambia domain between February-March 2024. When compared to Sumatra (Figure 5.6), the diurnal patterns show similarities, with initiation events producing a narrower signal that peaks earlier than intensification. However, both initiation and intensification events peak approximately 2 hours later over Zambia, as compared to Sumatra.

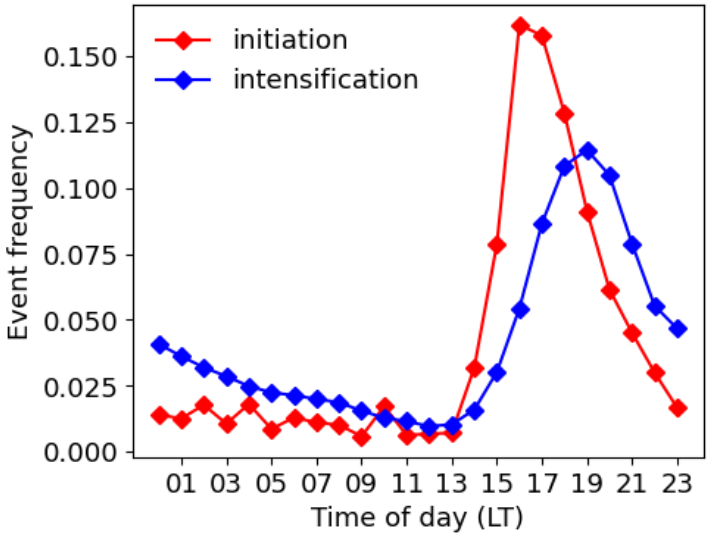


Figure 5.16: The frequency of observed events for convection intensification and convection initiation at each hour of the day over Zambia between February - March 2024.

Figure 5.17 presents a case study of the original SII-NowNet (trained over Sumatra) applied to Zambia (differing colour bar scales are used for intensification and initiation nowcasts as in Figure 5.8). The intensification and initiation nowcasts are valid at 13:00 LT on 15th March 2024 and only lead times within the limit of usefulness are presented. Figures 5.17a and e show a greater number of observed events for the Zambian domain case study, compared to the Sumatran domain case study (Figure 5.8). However, without any prior knowledge of this new domain, SII-NowNet still appears to show skill at capturing these widespread events. At 1 hours' lead time, the high intensification nowcast probabilities (Figure 5.17b) align with the bulk of observed events over the southern and northwestern regions of Zambia. Similarly, for the initiation nowcast (Figure 5.17f), SII-NowNet has captured a high proportion of the observed events e.g. over the east of Zambia. However, both nowcasts miss smaller areas of observed events too e.g. over the northwestern part of the domain in the intensification nowcast, and over the northeastern part of the domain in the initiation nowcast. At 2 hours' lead time

the intensification (Figure 5.17c) and initiation (Figure 5.17g) nowcasts produce more blurred images with less precision; however, the distribution of high probabilities still broadly follows those produced in the 1 hour lead time nowcast i.e. SII-NowNet continues to capture a high proportion of both events at 2 hours' lead time. For the 3 hour lead time intensification nowcast (Figure 5.17d), the high probabilities still align with some of the observed events e.g. western Zambia, but reduced skill at this lead time results in larger areas of observed events being missed e.g. north of Zambia.

To quantitatively understand how SII-NowNet performs over the Zambian domain, it was tested on 2,746 samples from March 2024. Moreover, to investigate whether knowledge of the domain would increase skill, SII-NowNet-Zam was tested over the same period. Figure 5.18 shows the AUC scores for initiation and intensification for both the original SII-NowNet and SII-NowNet-Zam. Although the original SII-NowNet has no knowledge of the Zambian domain, it still produces comparable AUC scores to those over the Sumatran domain (Figure 5.9a and b inset plot). For example, comparing SII-NowNet's 1 hour lead time AUC scores over Sumatra to those over Zambia, shows there is only a 2% and 9% reduction for intensification and initiation, respectively. Comparing the performance of SII-NowNet-Zam to the original SII-NowNet (applied to Zambia) shows mixed results. For initiation nowcasts at 1 hours' lead time, SII-NowNet-Zam increases the AUC score by 5%. However, at 2 hours' lead time, the original SII-NowNet version produces higher AUC scores than SII-NowNet-Zam (6% increase in AUC score over SII-NowNet-Zam). For nowcasting intensification at 1 – 2 hours' lead time, there is no consistent improvement in SII-NowNet-Zam over the original SII-NowNet, whereas at 3 hours' lead time the original SII-NowNet version outperforms SII-NowNet-Zam (4% increase in AUC score over SII-NowNet-Zam).

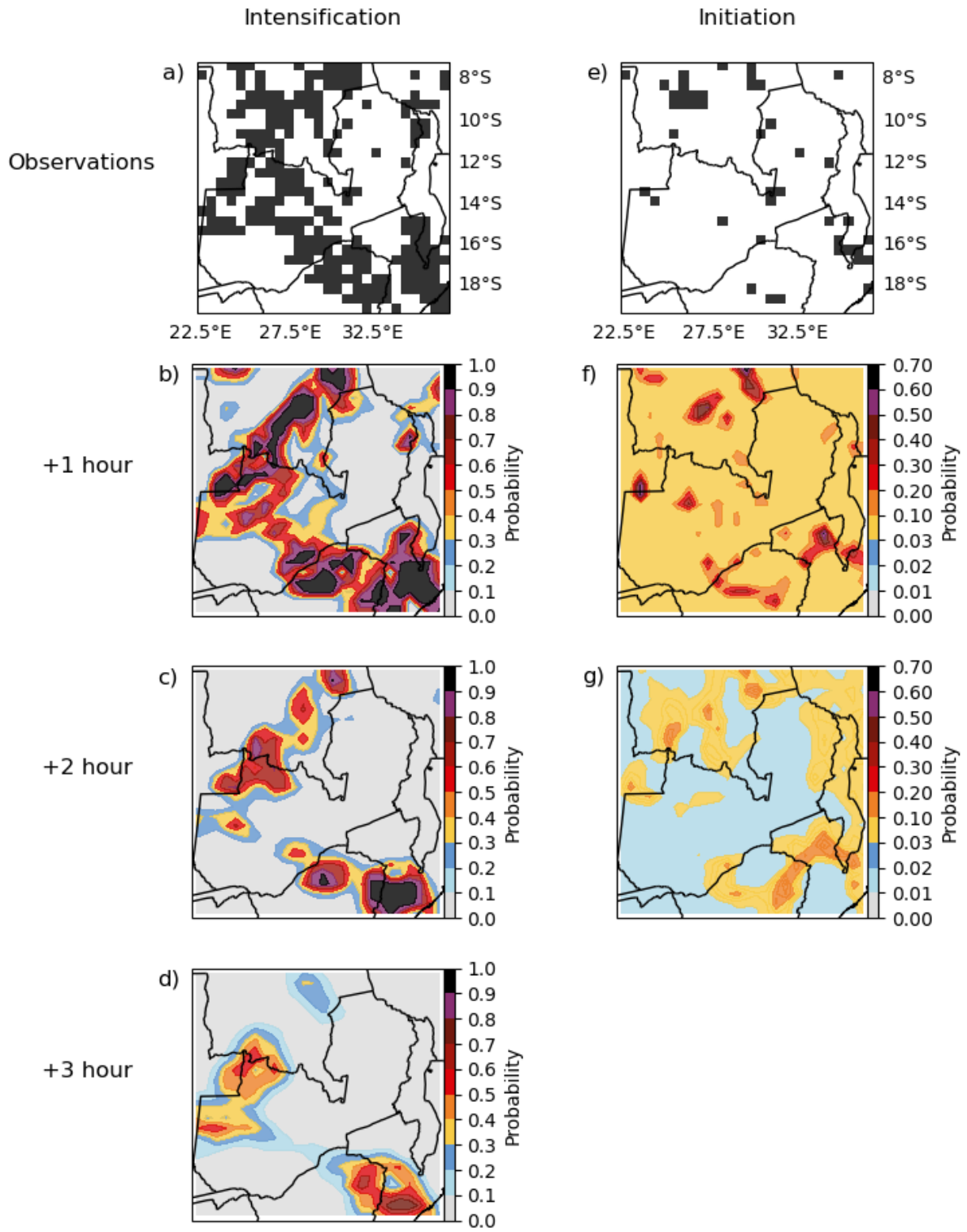


Figure 5.17: A case study of a) intensification and e) initiation events over Zambia at 13:00 LT on 15th March 2024. b-d) and f-g) show the probabilistic SII-NowNet nowcasts for intensification (1 – 3 hours' lead time) and initiation (1 – 2 hours' lead time), respectively. Each nowcast is valid at 13:00 LT. Different colour bar scales have been used for intensification and initiation nowcasts to account for their differing probability ranges.

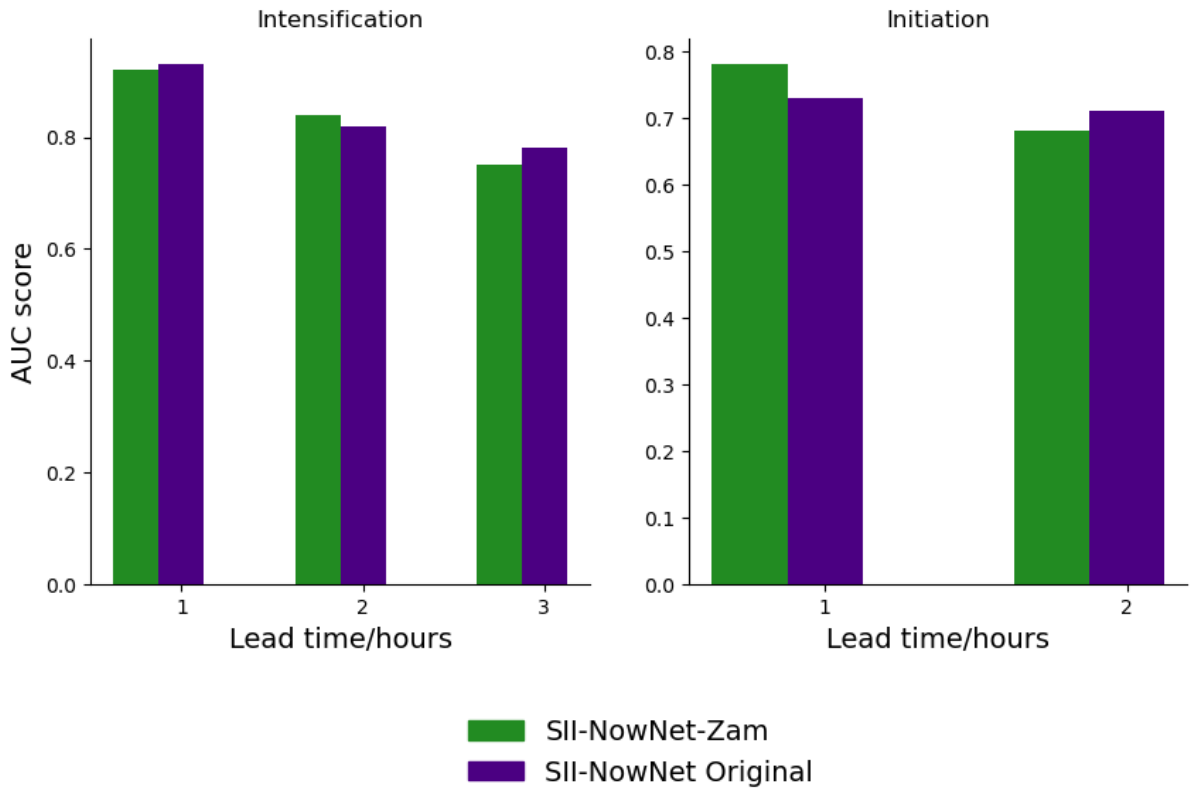


Figure 5.18: A comparison of AUC scores for SII-NowNet-Zam (with additional training over Zambia) and the original version of SII-NowNet (trained over Sumatra), both applied to the Zambia domain.

5.4 Conclusions and discussion

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

SII-NowNet performs best at nowcasting convection intensification. Over Sumatra, it outperforms both persistence and climatology across 1 – 6 hours' lead time, producing consistently higher likelihoods of capturing observed intensification events. SII-NowNet's nowcasts show reliability for up to approximately 3 hours' lead time, beyond which it begins to over-predict

the number of intensification events (producing false alarms) i.e. SII-NowNet's intensification nowcasts have a 3 hour lead time limit of usefulness.

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

Overall, SII-NowNet's results over Sumatra show that, by learning key patterns in the atmospheric conditions, it is able to discriminate between convective and non-convective events, proving its potential to support early weather warnings with multiple hours' lead time. This study also showed that SII-NowNet is able to capture geographical location in regard to the underlying controls of convection. Spatial feature heatmaps exemplified SII-NowNet's process of identifying specific locations of BT development to aid its predictions. Furthermore, comparing SII-NowNet's predictions on rotated input fields evidenced its ability to align high likelihoods to moving convective features. A more in depth comparison of the rotated predictions highlighted a tendency of the intensification nowcasts to produce more higher likelihoods over the land. This suggests that it has learned about the underlying surface conditions that control convection. This ability to capture the geographical location of convective controls is a key advantage of using a U-Net model, which is known to preserve spatial structure through convolutions and skip connections (Ronneberger et al. 2015). This ability to condition its output on both the input values and the spatial structure has been identified in some of the early application of U-Nets to spatial imagery (Long et al. 2015; Isola et al. 2017).

By nowcasting convection initiation and intensification in the Tropics using solely satellite data, SII-NowNet is a unique tool. It is therefore not appropriate to make a direct comparison to other studies, but instead use them as context for model performance. In Korea (subtropics), previous studies have presented nowcasting tools that produce an 82% (on a 6 km grid scale; Han et al. 2019) and 85% (on a 4 km grid scale; Lee et al. 2017) chance of capturing observed convective events (both studies used radar data for identifying convection occurence and produced deterministic models). For intensification events at a 1 hour lead time, SII-NowNet was thresholded to produce deterministic output with an 82% chance of capturing observed events.

Therefore, not only does SII-NowNet provide probabilistic information, but it can also produce similar deterministic results to previous studies. Mecikalski et al. (2021) used radar, satellite and lightning data to produce a probabilistic convection initiation detection model (i.e. 0 hour lead time) on a 10 km grid scale, which produced an AUC score of 0.81, comparable to the AUC score of 0.82 produced by 1 hour lead time SII-NowNet initiation nowcasts. Lagerquist et al. (2021) used U-Net models (trained on radar data) to produce probabilistic fields of convection on a $\tilde{1}6$ km scale over Taiwan and showed similar reliability results for 1 hour lead time nowcasts. However, for 2 hours' lead time, their results show that the U-Net over-predicted the likelihood of convection occuring (i.e. dropped in reliability), whereas SII-NowNet maintains its reliability for both intensification and initiation nowcasts at this lead time.

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

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

To highlight SII-NowNet's applicability to other regions of the Tropics, its performance was also evaluated over Zambia. Without any prior knowledge of the new domain (i.e. no re-training or region-specific tuning), SII-NowNet was able to produce skilful nowcasts of intensification and initiation, with comparable scores to those produced over Sumatra. The slight reduction in skill over Zambia is likely due to differences in the meteorology of the two regions. This was exemplified by their differing diurnal cycles of initiation and intensification. When additional training was applied to SII-NowNet on BT data over Zambia (to encompass domain knowledge),

there was no consistent improvement over the original version of SII-NowNet (trained over Sumatra). However, this may be due to the considerably shorter training period used for the additional training. Overall, the results from the Zambia domain are evidence that SII-NowNet can provide skill in other areas of the Tropics too.

Overall, SII-NowNet has shown that it can provide skilful probabilistic nowcasts of Tropical convection development in satellite observations. A key advantage of SII-NowNet is in its use of simple, freely available inputs, meaning that it can be applied to any region that is covered by a geostationary satellite. This versatility means that SII-NowNet has great potential for future use as an operational nowcasting tool.

Chapter 6

N-FOREWARNS: Regional nowcasting of surface water flooding with a 3-hour lead time predictability limit

6.1 Introduction

Since the Flood and Water Management Act 2010 (UK Parliament 2010), UK local government authorities have been responsible for the management of local SWF risk. However, varying levels of SWF exposure across the UK means that capability to undertake this management role differs between local authorities (Speight et al. 2025). To support local authorities in England and Wales, the Flood Forecasting Centre (FFC), a partnership between the Environment Agency and Met Office to improve flood risk guidance, issue a daily Flood Guidance Statement (Flood Forecasting Centre 2022). The Flood Guidance Statement provides a 5-day flood risk for surface water, river, groundwater and coastal flooding by using a risk matrix approach that details both likelihoods and impacts (Pilling 2016). The Flood Guidance Statement is primarily targeted at users involved in strategic and operational planning decisions for emergency response to flooding. A key source of information for the Flood Guidance Statement comes from the MOGREPS-UK NWP model (see section 3.2.2 for full description). This chapter focusses on substituting NWP

with nowcasting techniques for rapid, short-term prediction of SWF risk.

Flood fOREcasts for Surface WAter at a RegioNal Scale (FOREWARNS) was originally developed to provide national-scale surface water flood forecasts for whole calendar days, at 1 to 4 day lead times (Maybee et al. 2024). It generates reasonable worst case rainfall scenarios (Böing et al. 2020) from MOGREPS-UK forecasts and compares them to predefined look-up flood thresholds from offline hydrological modelling. Maybee et al. (2024) showed that FOREWARNS is a suitable tool for operational forecasting of surface water flooding. However, a key outcome of this study was that, given the challenges of forecasting rapidly intensifying convection with NWP, producing SWF predictions at shorter lead times (0-6 hours) using rainfall nowcasts, could fulfil the user-need for more reliable forecasts.

This chapter presents nowcasting-FOREWARNS (N-FOREWARNS), a modified version of FOREWARNS that is driven by probabilistic nowcast rainfall from STEPS and can produce maps of SWF risk predictions every hour with lead times of 1 to 6 hours. The aim of this development is to test the ability of increasing the latency and temporal resolution of the flood forecasts, whilst maintaining good forecast skill. This study presents the results of a SWF forecasting testbed, hosted by the Met Office and FFC during the 2024 summer season, in which expert participants provided feedback on the performance of N-FOREWARNS alongside current operational SWF forecast tools. Additionally, a quantitative verification of N-FOREWARNS is presented across lead times of 1 – 6 hours for 70 recorded SWF events from 2013 – 2022, and for 3,470 potential issuances across the 2024 summer period.

6.2 Data and methods

6.2.1 Rainfall radar data

For ground truth in the quantitative verification, this study uses ground-based radar retrievals of rainfall as input into SWF models (see section 6.2.3 for details on R-FOREWARNS). Rainfall data sets derived from 15 C-band radars across the UK are used to form the Met Office Nimrod system (Golding 1998; Harrison et al. 2012) which was used in this study. The output is produced on a 1 km grid every 5 minutes, and has been verified against rain gauge measurements across the UK (Harrison et al. 2012).

6.2.2 SWF impact reports

Identifying SWF events and their observed impacts presents a key challenge in SWF prediction and verification. Currently, the most effective approach at national scale is to collate information using algorithms that scan a variety of sources such as social media, news publications and official reports, which mention surface water flooding (De Bruijn et al. 2019; FloodTags 2025).

During the testbed part of this study, Hydrometeorologists at the FFC collated reports of SWF events across the UK, which provided a source of observational data for verification. Reported SWF events were identified from social media and professional partner sources. The location, dates and impact severity of each reported event are identified and then the data is interpolated onto a county-scale map (e.g. Figure 6.1).

Figure 6.1: *Left* An example of a county scale map of reported SWF, with their associated severity level, for 15th July 2024 (impacts may have been reported at any time during the day). *Left* An image extracted from a post on https://x.com/ giving evidence of minor surface water flooding in Worthing, Sussex. This SWF event report is scaled up to the county level to produce the mapped impact level shown (left).

Maybee et al. (2024) built a record of observed minor and significant impact SWF events for Northern England using the Global Flood Monitor (GFM; https://www.globalfloodmonitor.org/, last access: 06 March 2025) to identify observed events between 2013 – 2022 May – October. In order to make a comparison, this study chooses 70 SWF event days from that dataset (chosen based on data availability) as part of the quantitative verification of N-FOREWARNS. A verification over the full summer 2024 period is also presented to understand N-FOREWARNS's performance on a more climatologically representative dataset.

Applying methods that scan various sources for SWF events enables a range of useful impact information to be captured. However, SWF events that are not reported through these sources will be missed (De Bruijn et al. 2019) e.g. events with localised impacts. Acknowledging this key caveat when interpreting any impact-based SWF verification is important (Maybee et al. 2024).

6.2.3 Surface Water Flood Forecast Tools

Nowcasting-Flood fOREcasts for Surface WAter at a RegioNal Scale (N-FOREWARNS)

N-FOREWARNS is a modified version of FOREWARNS (Maybee et al., 2024). FOREWARNS was co-designed with users (Birch et al., 2021) and is intended to be used by professional forecasters in the routine production of national guidance such as the daily Flood Guidance Statement by the FFC. The FOREWARNS method has two key steps:

1. It generates reasonable worst-case rainfall scenarios (RWCRSs) for a range of rainfall accumulation periods (1, 3 and/or 6 hours) across a 24-hour forecast period. This is done by applying the Böing et al. (2020) neighbourhood processing method to ensemble rainfall NWP forecasts. In this method each grid box is analysed along with the surrounding 30 km area. The maximum rainfall values within this neighbourhood are ranked, and the 98th percentile is computed to estimate extreme rainfall. Overall, this identifies the areas that are most at risk of heavy rainfall (as predicted by the forecast).

2. Rainfall accumulations from the RWCRSs are then translated into catchment-level SWF return periods, by conducting look-up comparisons against pre-run rainfall modelling (Vesuviano 2022) underpinning the UK Risk of Flooding from Surface Water maps (a dataset of national street-level SWF risk available at (https://check-long-term-flood-risk.service.gov.uk/map). The outputs from the look-up comparison of each scenario are then combined to create a single, reasonable worst case scenario SWF risk map of flood return periods (5, 10, 30, 100 and 1000 years).

N-FOREWARNS is run with three key modifications:

I. The NWP ensemble rainfall forecasts used in stage 1 of the FOREWARNS setup are replaced with STEPS nowcast rainfall predictions.

II. A 3-hour period of STEPS rainfall data is used as input instead of the 24-hour period

used in FOREWARNS. Therefore, N-FOREWARNS calculates RWCRSs using 1 and 3 hour accumulations periods only (unlike 1, 3 and 6 hours for FOREWARNS). A 3-hour input period was chosen as a balance between capturing the majority of past rainfall that causes a SWF event and the need for short lead times. The 3-hour period is selected so that, for a given SWF valid time, the period ends at the valid time. For example, for a N-FOREWARNS prediction of a SWF at valid time 1600 UTC, STEPS rainfall over the 3 hour period 1300 – 1600 UTC would be used as input.

III. The catchment-level output is linearly interpolated onto a 20 km grid for direct comparison against other SWF nowcasting tools, whereas the output was presented at the catchment scale (30-40km) in FOREWARNS.

As discussed in section 6.2.2, the lack of comprehensive observed SWF datasets presents a key challenge in evaluating SWF nowcast output directly against SWF impacts. Therefore, to enable a robust quantitative verification, a radar-driven version of N-FOREWARNS – referred to as R-FOREWARNS – is produced to act as a proxy for SWF ground truth. To do this, the STEPS rainfall data is replaced by the Nimrod (radar) rainfall observations from the same 3-hour period (as in step II). Figure 6.2 provides an example of N-FOREWARNS and the corresponding R-FOREWARNS. Defining R-FOREWARNS as the ground truth allows a direct evaluation of N-FOREWARNS. However, this comparison is an assessment on the STEPS rainfall input (and its ability to capture SWF risk) only and does not provide insight into N-FOREWARNS's capability in capturing recorded SWF events.

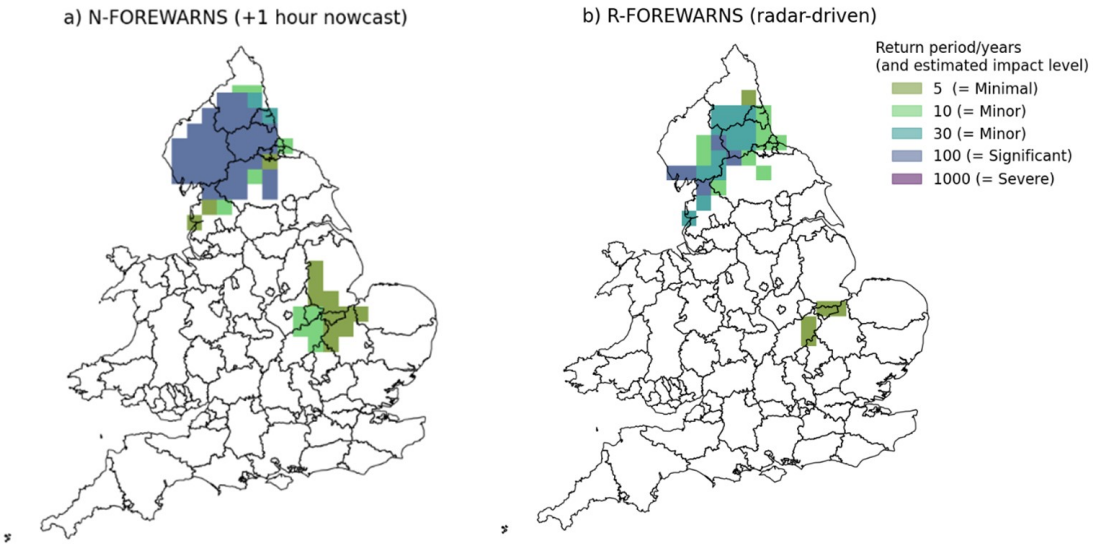


Figure 6.2: An example from 10th September 2023 valid at 15:00 UTC of a SWF warning from a) N-FOREWARNS (nowcast-driven with a 1 hour lead time) i.e. the prediction and b) R-FOREWARNS (radar-driven) i.e. the verification.

The 3-hour periods in N-FOREWARNS for step II above are only available from a single STEPS run for the 3, 4, 5 and 6 hour lead times (i.e. 0–3 hours, 1–4 hours, 2–5 hours and 3–6 hours from the STEPS initialisation). In order to extract a 3-hour rainfall period for the 1 and 2 hour lead time, data is extracted from previously initialised STEPS runs as shown in Figure 6.3.

Surface Water Flooding Hazard Impact Model (SWFHIM)

SWFHIM is a SWF forecast tool that is currently in operational use by the FFC (Aldridge et al. 2020) for lead times of ∼6-hours to 3 days. It produces probabilistic output on a 1 km grid scale for minimal, minor, significant and severe SWF impact levels (Figure 6.4). The SWFHIM method is split into three parts: hazard model, impact model and risk mapping. The hazard model component comprises the FFC's Grid-to-Grid hydrological model (Cole and Moore 2009), which is driven by ensemble NWP rainfall forecasts from MOGREPS-UK to produce a 1 km gridded surface runoff field. The impact model is based on the Impact Library - a static dataset describing predefined impact scenarios in terms of their trigger criteria and their associated impact levels (dependent on transport, property, population and infrastructure). The Impact Library is mapped onto a 1 km grid, enabling the impact model to translate hazard model output to impact levels at this scale. The hazard and impact models both operate on the individual ensemble members of the NWP rainfall forecast. In the risk mapping component, the outputs from the impact model and the hazard model are combined and upscaled to produce

ensemble-based likelihood maps for each SWF impact level (minimal, minor, significant and severe).

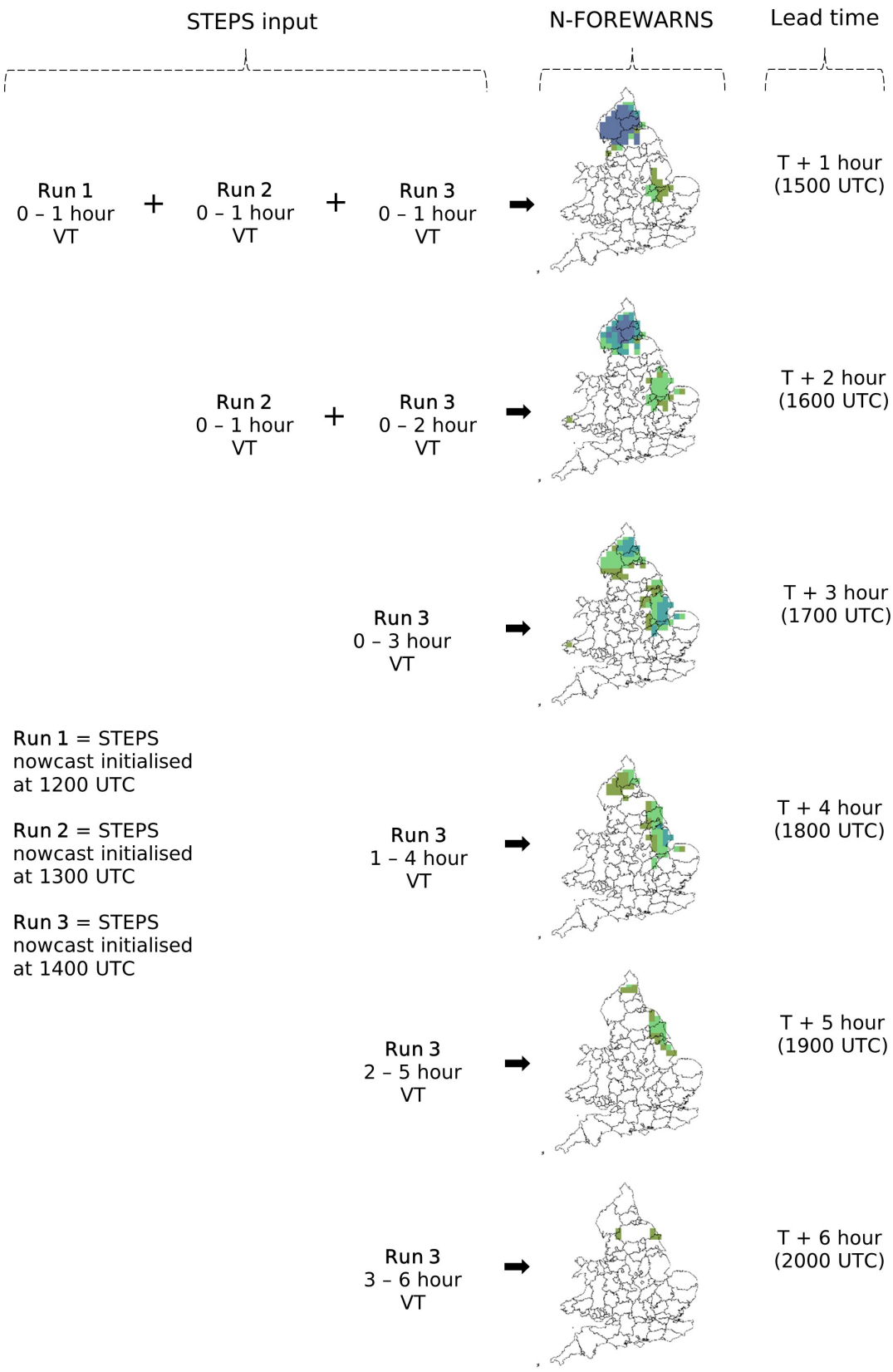


Figure 6.3: A diagram showing how the 3-hour period of STEPS rainfall data is inputted for each lead time (represented by each row). The STEPS run and period of valid time (VT) is shown for each N-FOREWARNS lead time output. An example from 10th September 2023 is show, which has been initialised at 14:00 UTC.

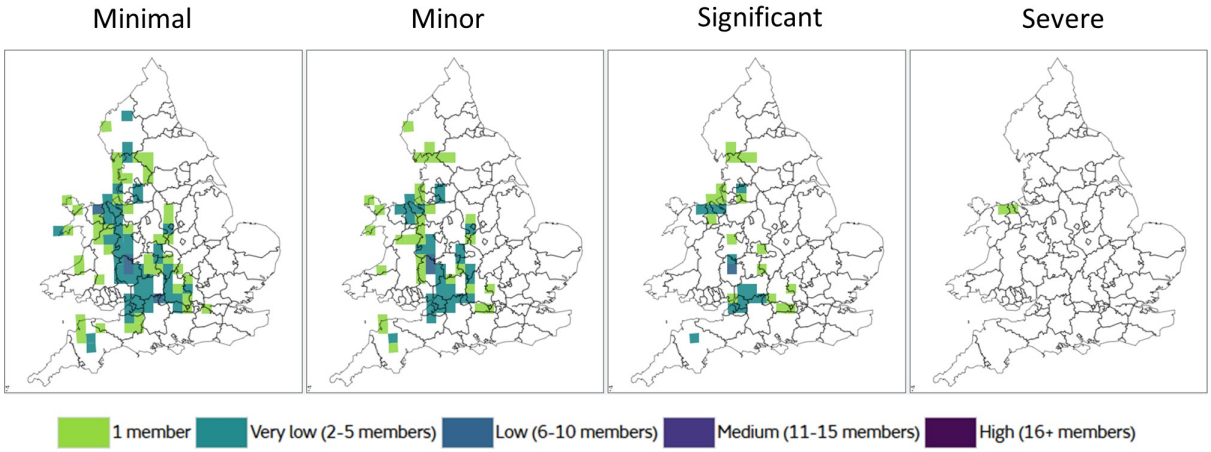


Figure 6.4: An example of a SWFHIM SWF forecast valid at 19:00 UTC on 12th May 2024.

For this study, SWFHIM was also run in nowcast mode using 3-hour periods. As with the modifications to N-FOREWARNS outlined earlier in this section, 3-hour periods of STEPS ensemble predictions were used. Outputs were upscaled to a 20 km grid (Figure 6.4). Table 6.1 provides the key comparison of N-FOREWARNS and SWFHIM.

Feature	**N-FOREWARNS**	**SWFHIM**
Nowcast rainfall field	STEPS	STEPS
Testbed output grid scale	20 km	20 km
SWF risk measure	Return period	Ensemble member count
Output format	Single reasonable worst-case scenario map	Impact map for each severity level

Table 6.1: Comparison between N-FOREWARNS and SWFHIM

Rapid Flood Guidance (RFG)

The FFCs RFG operational trial ran from 14th May to 30th September 2024 (Flood Forecasting Centre 2024) to support emergency responders to prepare for rapidly developing SWF events that occur due to summertime convection over England and Wales. The key aims of the RFG service were:

- To give short notice updates for England and Wales to supplement the daily Flood Guidance Statement.

- To be designed for responders who need to make decisions at a timescale of 0 to 6 hours.

During the trial, operational forecasters at the Met Office were responsible for issuing the RFG. Traditional observation sources such as composite and single-site radar parameters, rain gauges, satellite imagery and radiosonde ascents were analysed in combination with high resolution short-range NWP (MOGREPS-UK) and forecaster experience to generate the RFG output. Note, N-FOREWARNS and SWFHIM in nowcasting mode were not used in the RFG production process. Each RFG contained a forecaster-drawn polygon over the region of impact, along with written information describing the timing, likelihood and characteristics of the weather that is driving the rapid flooding (Figure 6.5).

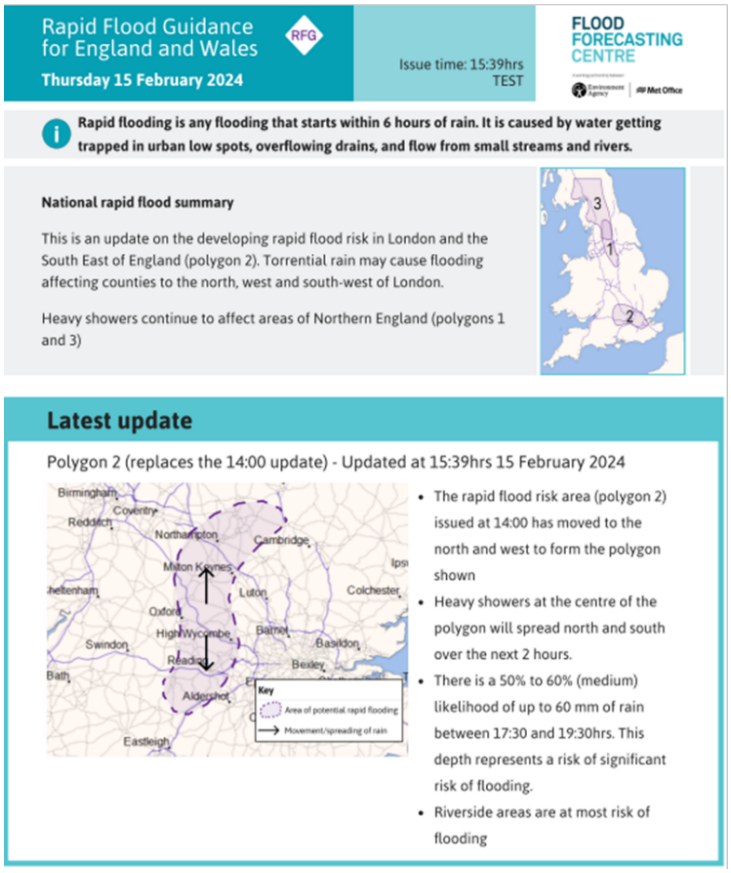


Figure 6.5: An example of a Rapid Flood Guidance issued at 15:39 UTC on 15th February 2024.

6.2.4 UK summer 2024 SWF testbed

Testbed summary and aims

Aligned in time with the RFG operational trial (summer 2024), the FFC hosted a SWF nowcasting testbed in collaboration with researchers from the University of Leeds and the University of Oxford. During the testbed, the outputs from N-FOREWARNS and SWFHIM in nowcasting mode were evaluated by meteorological/hydrological expert participants and compared to the RFG and to observations of SWF impacts. The key aims of the testbed were to:

1. Compare and evaluate the performance of N-FOREWARNS and SWFHIM SWF nowcasts.
2. Assess whether the use of N-FOREWARNS and/or SWFHIM nowcasts by operational forecasters would have added value to the production of the RFGs.

Testbed structure

During the testbed, 20 events were analysed covering 11 different days (see Table A.1 in supplementary for information on all events) with each event aligning with the issuance of an RFG warning. N-FOREWARNS and SWFHIM were then run at the same initialisation and lead time to enable a direct comparison between all three tools. A total of 50 individual participants took part across the duration of the testbed. For each event 5 – 20 participants (including the organisers of the study) were split into groups, with each group providing an independent subjective assessment of the N-FOREWARNS, SWFHIM and RFG predictions. Across all events a total of 92 groups provided subjective assessments (i.e. 92 individual data points). Participants came from a range of background expertise: hydrometeorologists, operational forecasters and university academics. The session focussed on a post-event assessment and was structured in the following way:

1. The flooding event (a recent event, if possible, from the previous week) was introduced to the participants with an overview of the meteorological conditions.
2. Participants groups were provided with the SWF nowcasts from N-FOREWARNS and SWFHIM and, if appropriate, asked to draw a flooding impact polygon over England and Wales based on each forecast (Figure 6.6). They were then required to identify the most appropriate impact severity for each polygon from Minor, Significant and Severe.

3. The rainfall radar observations, R-FOREWARNS and the observed impacts (Figure 6.7), plus the issued RFG were revealed to the groups to compare and subjectively verify. The groups were then asked to complete a survey (Table A.2), which focussed on evaluating different aspects of the performance of N-FOREWARNS, SWFHIM and RFG.

4. Finally, each group was asked to feedback their key points during an all-participant discussion.

Drawing the impact polygons in step 2 meant that participants were forced to interpret and consider strengths and weaknesses of the N-FOREWARNS and SWFHIM tools. Furthermore, it provided benchmarks to compare with the RFG polygon.

6.2.5 Verification metrics

During the testbed the participants were asked to qualitatively categorise the SWF nowcasts that were produced by N-FOREWARNS and SWFHIM into one of: hit (SWF nowcast event aligns with the SWF observed event), miss (the SWF observed event is not captured by the SWF nowcast), correct rejection (no SWF event is produced in the nowcast or observation) and false alarm (a SWF event is nowcasted but no SWF event is observed). These results are presented in section 6.3.1.

For the quantitative analysis presented in section 6.3.2, the POD, FAR and SR skill scores (see section 3.2 for full description) are derived from the verification of N-FOREWARNS against R-FOREWARNS.

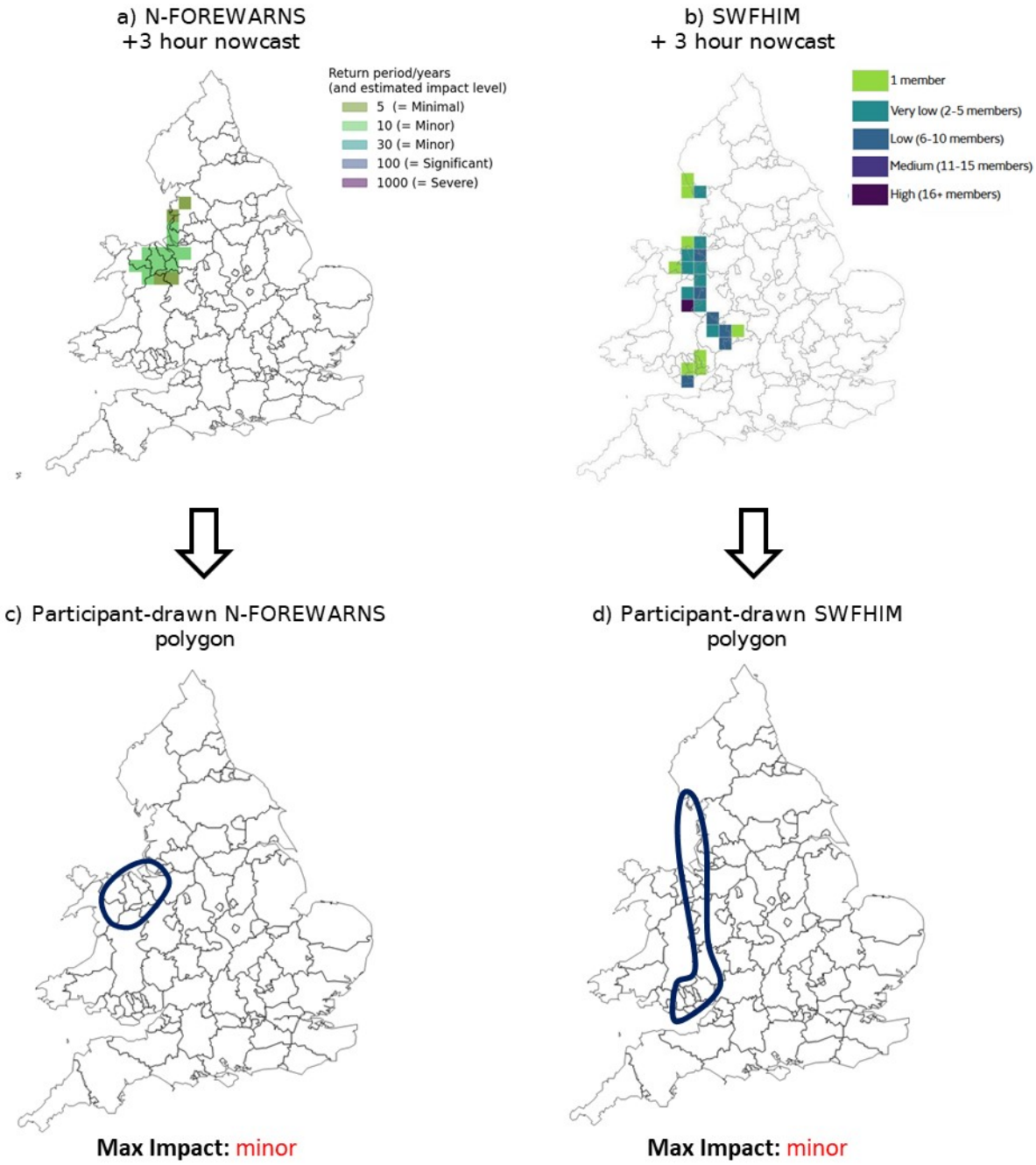


Figure 6.6: Using the N-FOREWARNS and SWFHIM SWF nowcasts (a and b; only minor impact level shown here for SWFHIM example), participants drew SWF warning polygons (c and d) and associated them with a maximum impact. This example is taken from an event on 12th May 2024. The SWF nowcasts have a lead time of 3 hours and are valid at 18:00 UTC.

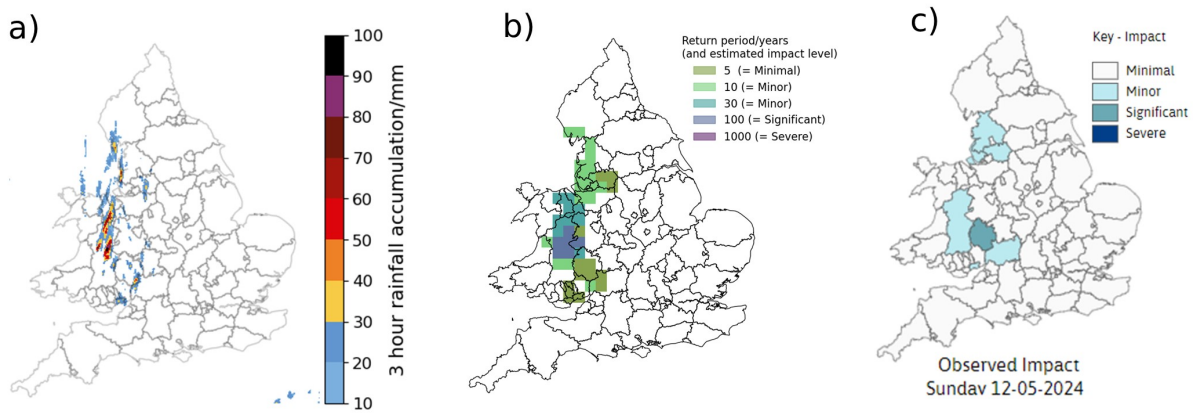


Figure 6.7: An example (valid at 18:00 UTC on 12th May 2025) of the observation sets that were revealed to the participant groups after they completed the polygon drawing exercise (step 2). They provide the a) 3-hour radar rainfall accumulation taken from the 3 hours prior to the valid time, b) corresponding R-FOREWARNS output at the valid time and c) the reported SWF impacts collected by the hyrdrometeorologists at the FFC.

6.3 Results

6.3.1 Testbed results

Testbed case study events

Table 6.2 contains five example events selected from the 20 covered in the testbed. These five events were subjectively chosen as they capture a range of different weather scenarios and SWF nowcast outputs, enabling the key advantages and disadvantages of each tool to be discussed (only minor impacts presented for SWFHIM output).

During Event 1, rainfall accumulations of approximately 20 mm were observed in the north of England, but no observed flood impacts were reported. N-FOREWARNS produced a minimal prediction that aligned with the rainfall accumulation, whereas the SWFHIM minor impact prediction was more widespread with lower precision, covering the majority of northern England and reaching south to Norfolk. The RFG contained a polygon constrained to the northeast, covering the rainfall regions, but also extending into regions where no impactful rain was observed.

In event 2 SWFHIM produced a more constrained and coherent minor impact prediction. Its low – medium likelihood regions aligned closely with the accumulated rainfall and coincided with the regions where minor and significant observed impacts were reported. N-FOREWARNS, on

the other hand, produced minor and minimal impact prediction over northern Wales, missing the majority of impacts further south. In contrast to both SWFHIM and N-FOREWARNS, the RFG warning was much larger and widespread across Wales, covering both affected and unaffected regions.

Event 3 is focused in the southeast of England (circled in red), with observed impacts reported over London and Sussex. All three SWF tools produced predictions over the southeast, with N-FOREWARNS and SWFHIM being more tightly constrained to London. Again, the RFG warning polygon was much larger and more widespread, containing regions where no rainfall or impacts are observed. Isolated, low likelihood predictions in Cumbria and Lincolnshire exemplify the low precision nature of SWFHIM, whereas N-FOREWARNS is larger but more focused over the southeast region.

One of the most impactful flooding events of the testbed was Event 4 on 1 August, which caused significant impacts to businesses, buildings and transport networks in Winchester, Hampshire. 50 – 60 mm of rainfall accumulated between 15 UTC – 18 UTC, with minor impacts also observed in Berkshire and Surrey. N-FOREWARNS produced a widespread significant impact prediction over the region, whilst SWFHIM produced minor impact predictions that were more focussed over the surrounding regions of Hampshire. The RFG warning, on the other hand, produced a polygon over Sussex, which did capture some isolated rainfall there, but missed the bulk of the convective rainfall and impacts further west.

During event 5, localised convective rainfall over northeast of Hampshire and the Isle of Wight produced SWF events that were categorised as minor (circled in red). This is a typical example of N-FOREWARNS under-representing the flood risk for a minor observed event, when the rainfall is particularly isolated. SWFHIM approximately identifies the correct area, producing a minor prediction just east of Hampshire. However, it also produces a minor prediction over central Wales, where no rainfall is observed, and no SWF events are reported. The RFG polygon is more constrained to Hampshire, capturing the rainfall and impacts well.

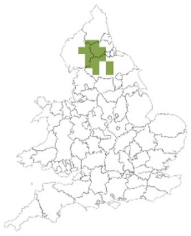
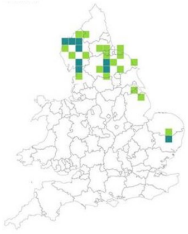
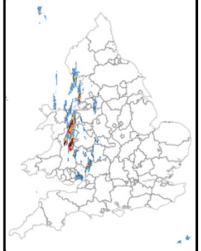
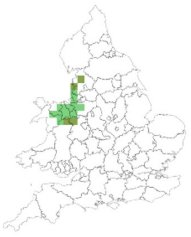
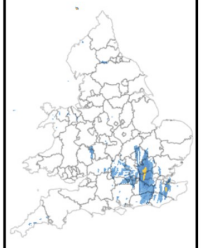
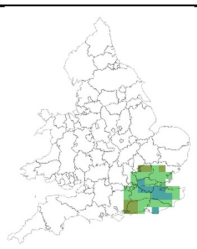
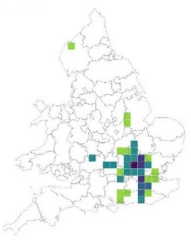
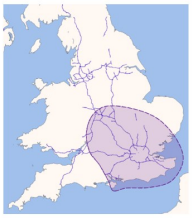
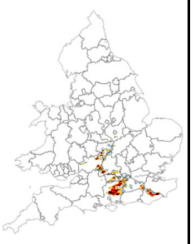
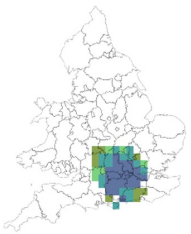
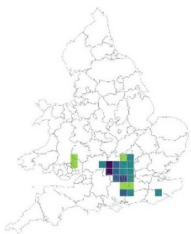
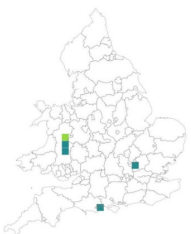
	Observed Impact Minimal Minor Significant Severe	3 hour radar accumulations (>10mm) 10 20 30 40 50 60 70 80 90 100 3 hour rainfall accumulation/mm	N-FOREWARNS Return period/years 5 (= Minimal) 10 (= Minor) 30 (= Minor) 100 (= Significant) 1000 (= Severe)	SWFHIM (minor) 1 member Very low (2-5 members) Low (6-10 members) Medium (11-15 members) High (16+ members)	RFG (warning area enclosed by polygon)
Event 1 15th Jun 2024 Initialised: 1300 UTC Valid until: 1600 UTC (+3 hour lead time)					
Event 2 12th May 2024 Initialised: 1500 UTC Valid until: 1800 UTC (+3 hour lead time)					
Event 3 15th Jul 2024 Initialised: 1800 UTC Valid until: 2000 UTC (+2 hour lead time)					
Event 4 1st Aug 2024 Initialised: 1600 UTC Valid until: 1800 UTC (+2 hour lead time)					
Event 5 8th Sep 2024 Initialised: 0900 UTC Valid until: 1100 UTC (+2 hour lead time)					

Table 6.2: Information on the 5 selected events from the testbed. For each event (row), the reported SWF impacts, 3-hour radar accumulation, N-FOREWARNS warning, SWFHIM minor impact warning and RFG output are shown. For each event, the SWF prediction tools were initialised with a lead time to match the RFG (ranging from 1 – 3 hours lead time). For days when multiple observed impacts were reported, the red circles identify those that occurred during the valid time.

These examples capture the key behaviours of each SWF prediction tool. N-FOREWARNS generally produces fewer predictions of SWF than SWFHIM and RFG (e.g. Events 1, 2 and 5), which for low rainfall accumulation events (e.g. Event 1) enables it to correctly suppress the prediction severity. However, for higher, localised rainfall amounts that cause minor events (e.g. Event 5), it can fail to capture potential SWF events. When N-FOREWARNS does identify high likelihood of a SWF event (during higher rainfall accumulation events e.g. Events 3 and 4), it is often in the correct location. SWFHIM on the other hand is more sensitive to the rainfall input and will produce a SWF prediction for the majority of events (e.g. Events 1, 2, 3, 4 and 5), capturing minor impact events more often than N-FOREWARNS (e.g. Event 1). However, due to this sensitivity, SWFHIM's output has low spatial precision that can make it difficult to interpret (e.g. Events 1, 2, 3 and 4). Furthermore, this can often make SWFHIM predictions appear too spatially widespread (e.g. Event 1). In contrast, the RFG warning, which is forecaster-drawn, is always a coherent polygon and mostly captures the events (e.g. Events 2, 3, 4), but is often too widespread, meaning unaffected regions are included in the flood warning.

A comparison of the R-FOREWARNS output (Table A.3) with the corresponding reported SWF impact observation for every testbed event shows that the SWF prediction in R-FOREWARNS covers the location for approximately 75% of the reported events (a more in depth verification against R-FOREWARNS is presented in section 6.3.2). Although this result is derived from testing over only 20 events, it suggests that, when accurate rainfall data is used (i.e. radar data), the N-FOREWARNS methodology can capture the majority of SWF events. Furthermore, it suggests that the main uncertainties come from the rainfall data that is input into N-FOREWARNS, rather than it's processing into a SWF prediction.

Testbed survey responses

Figures 6.8-6.10 show the results from three survey questions that participant groups were asked during the testbed for all 20 events (92 subjective assessments in total). These questions aimed to cover participant opinions on three key areas of the SWF nowcast tools: location, severity and overall performance assessment.

In terms of impact severity, the most common participant response was that the participant-drawn SWF polygons (see section 6.2.4) of N-FOREWARNS and SWFHIM were 'about right'

(47% and 39% respectively, Figure 6.8a and b). However, for both SWFHIM and RFG, a large proportion (49% and 64%, respectively) believed that the polygons had impact severities that were 'Slightly too high' or 'Much too high' (Figure 6.8b and c). Across the testbed SWFHIM produced low likelihoods (1 – 5 members) of significant impacts for the majority of events, when only 5 observed significant events were reported (Table A.1). For N-FOREWARNS the second highest impact severity category was 'Miss', with only 23% of participants thinking that it was too high or too low. This suggests that, although N-FOREWARNS often doesn't produce enough predictions, when it does, they are likely to have the correct severity. Both N-FOREWARNS and SWFHIM show a wide distribution of severity scores, reflecting a range in participant opinions on their performance (Figure 6.8d). The RFG, however, produces a much narrower distribution with a mean score of 2 i.e. its tendency to over-predict on severity levels is more consistent than N-FOREWARNS and SWFHIM.

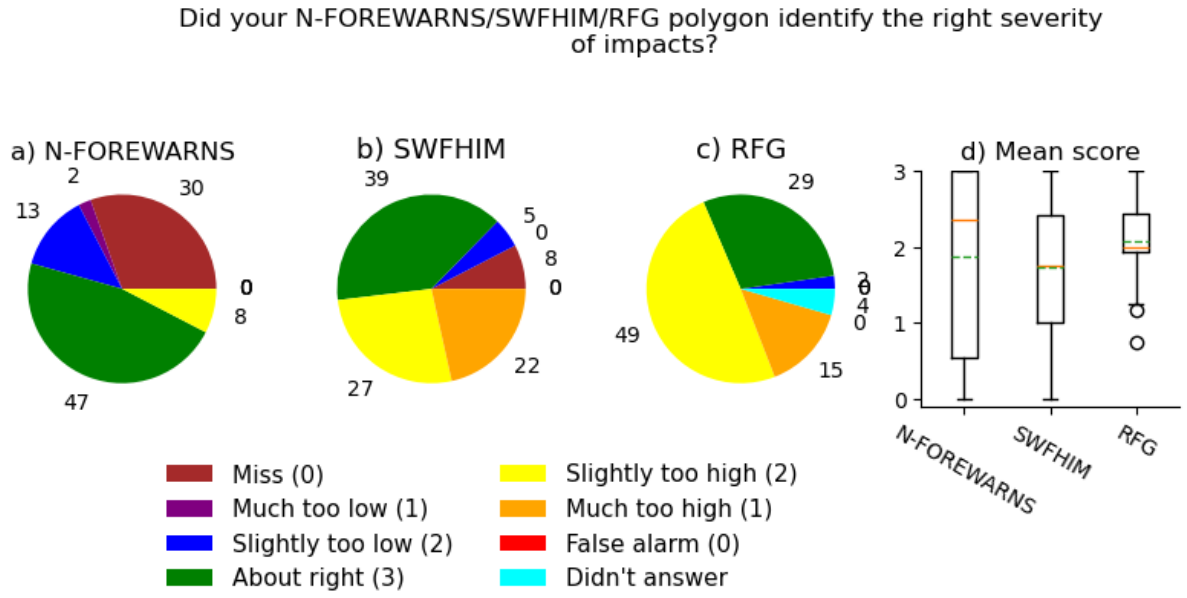


Figure 6.8: Across all events, the percentage of participants who voted for each severity level evaluation category for a) N-FOREWARNS b) SWFHIM and c) RFG. Each category has an associated score (shown in brackets), with a higher score representing a better performance. d) Shows the distribution of these scores, with median (orange) and mean (green) scores, for each SWF prediction tool across all events.

When considering participants' opinions on polygon location, N-FOREWARNS produced the highest proportion (34%) of 'Miss' category (Figure 6.9a). This is exemplified in Event 2 and Event 5 in Table 6.2, when the nowcast does not capture the spatial distribution of the

impacts. However, only 7% of participants reported that the N-FOREWARNS polygon was in the 'Incorrect location', suggesting that when N-FOREWARNS does produce a prediction, it is often considered to be in the correct location (as also seen from 33% of participants voting for the 'About right category'). The reported participant opinions of SWFHIM (Figure 6.9b), on the other hand, oppose that of N-FOREWARNS, with far fewer participants voting SWFHIM's location as a 'Miss'. The largest proportion of SWFHIM location votes was 'About right', however, 26% agreed that its polygon was in an 'Incorrect location'. As demonstrated in Table 6.2, SWFHIM will produce predictions in approximately the right location, resulting in a hit; however, predictions often have low spatial precision, meaning flood warnings are given to regions that are unaffected, resulting in false alarms. Similar to SWFHIM, a large proportion of participants (29%) reported that the RFG polygon locations were 'About right' (Figure 6.9c). However, the strongest participant response was that the RFG polygons were too large ('Slightly too large' and 'Much too large' categories combined to 47%). This can clearly be seen throughout Table 6.2 where the RFG often produces large, widespread polygons (Events 2, 3 and 5 in Table 6.2). The distribution of scores for location (Figure 6.9c) show a large spread of participant opinions for polygon location for N-FOREWARNS, SWFHIM and RFG. Again, this reflects the ranging opinions of participants and the tools' inconsistent ability to predict the location across a range of event types.

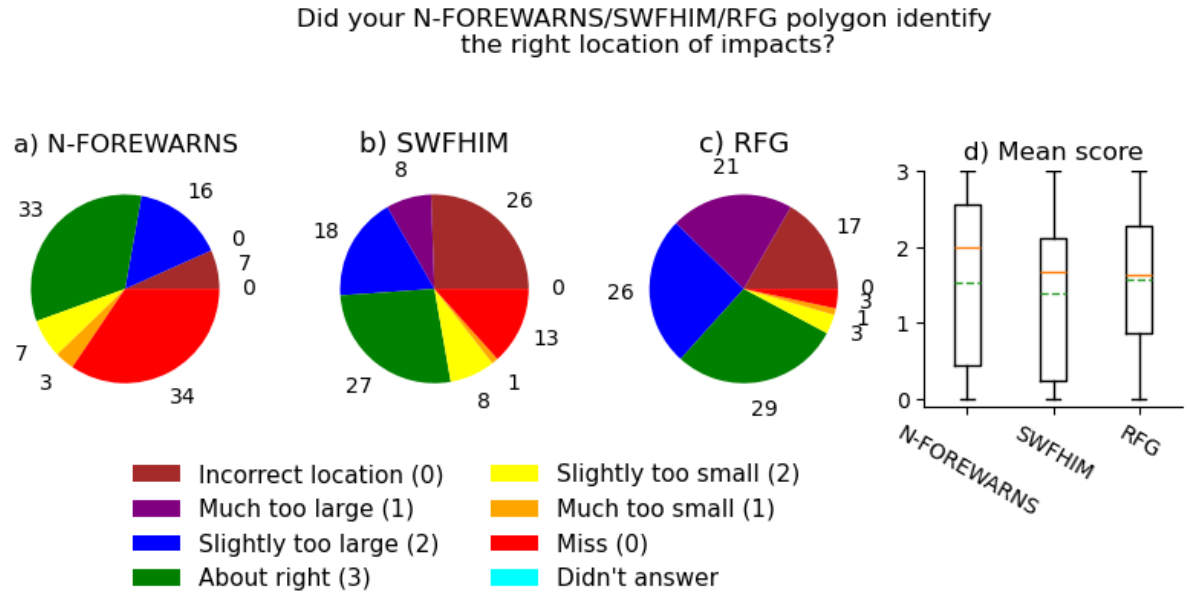


Figure 6.9: Same as Figure 6.8 but for location evaluation categories.

Figure 6.10 shows the participants' responses when asked to give an overall assessment of N-FOREWARNS and SWFHIM in terms of a hit, miss, false alarm or correct rejection. SWFHIM produces a greater proportion of votes for a hit and a false alarm, whereas N-FOREWARNS produces a greater proportion of votes for correct rejection and miss – a reflection of the tools behaving in opposing ways. SWFHIM produces a greater number of SWF predictions, some of which are correct – generating a higher hit score – but at the same time generating more false alarms. N-FOREWARNS on the other hand produces far less SWF predictions, resulting in a greater number of correct rejections, but also producing more misses.

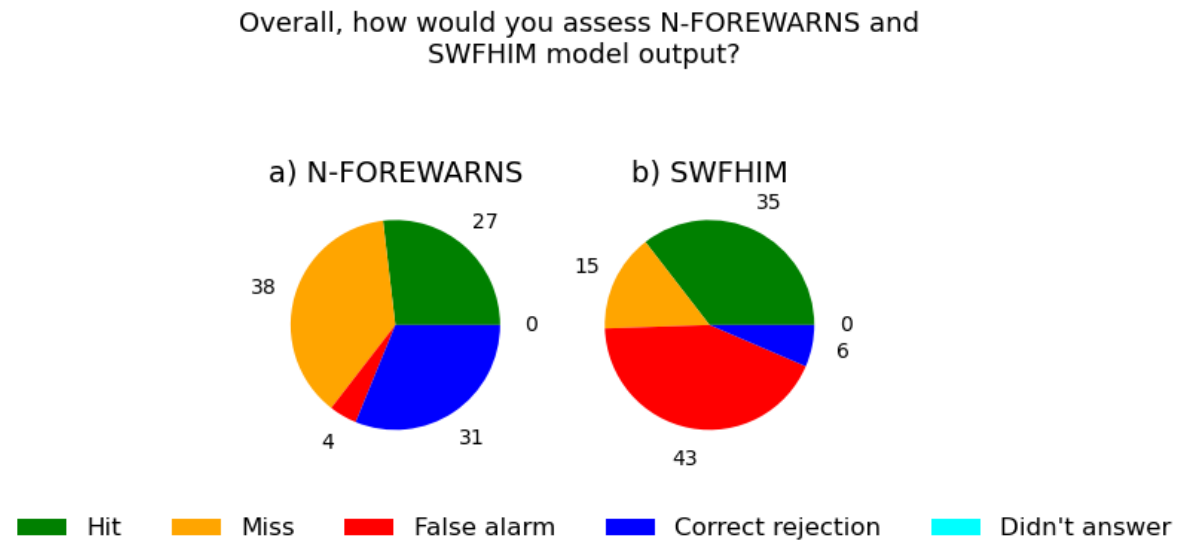


Figure 6.10: Across all events, the percentage of participants who voted for each overall evaluation category for a) N-FOREWARNS and b) SWFHIM.

6.3.2 Quantitative N-FOREWARNS analysis

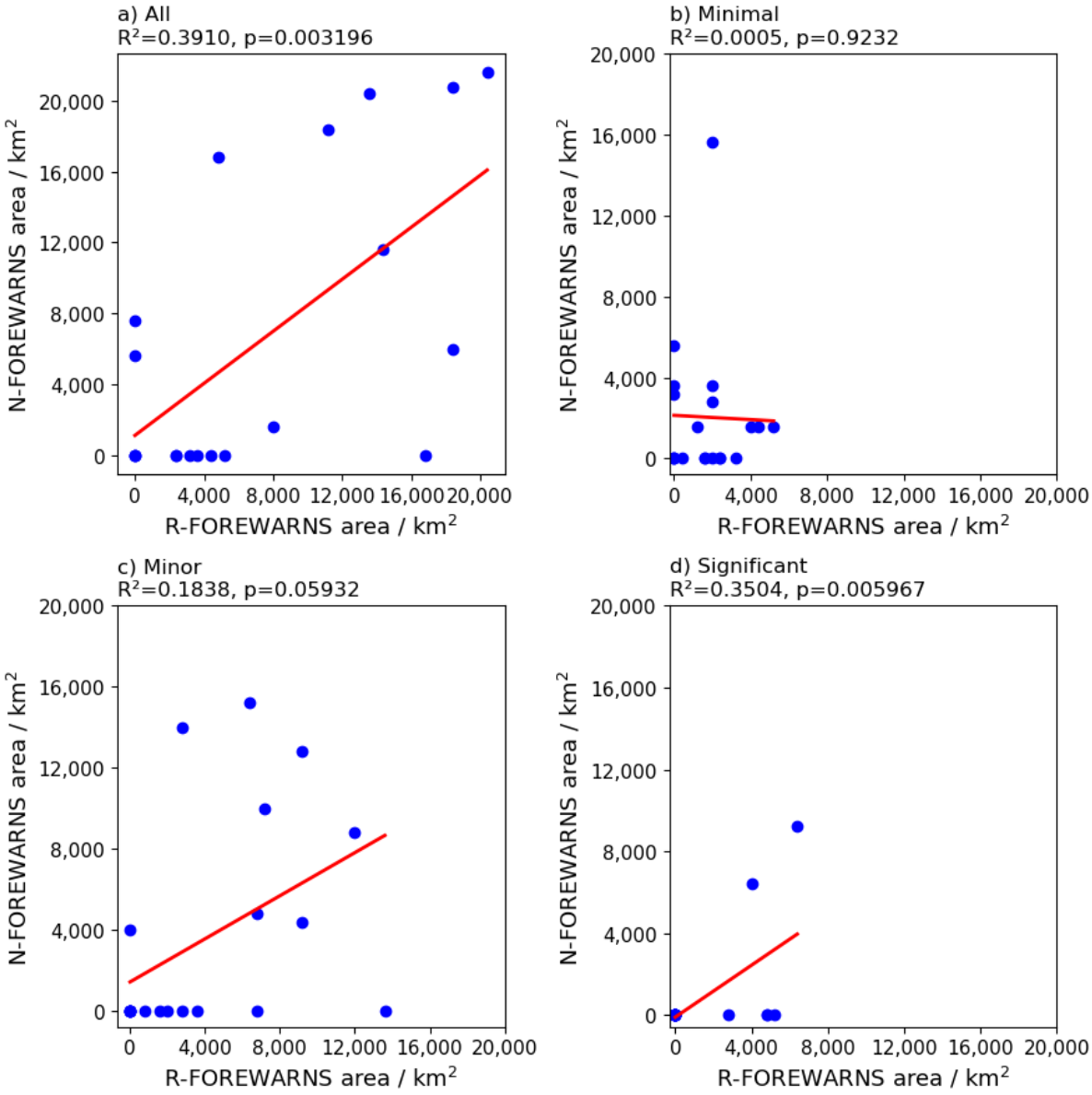


Figure 6.11: A comparison of the warning areas produced by N-FOREWARNS and R-FOREWARNS for a) all severity level events, b) minimal severity level event, c) minor severity level events and d) significant severity level events. A linear regression is fitted to each set of data, along with the R^2.

To verify the testbed participants' feedback on N-FOREWARNS's nowcasts, Figure 6.11 compares the warning areas produced by N-FOREWARNS and R-FOREWARNS for each severity level, for every testbed event. For each category a linear regression (red line) has been fitted to the data - a STEPS rainfall field that perfectly matches the radar rainfall would result in all points lying along x=y. Across all categories there is a weak correlation between the

N-FOREWARNS and R-FOREWARNS warning areas (all R^2 values below 0.4) and only the 'All' and 'Significant' categories have statistical significance ($p<0.05$). One of the key participant responses was that N-FOREWARNS produced too many misses (Figure 6.9) and only 16% reported that predicted impacts were too large. However, Figure 6.11a shows that N-FOREWARNS produces a more even spread of warnings that are too large or too small (or misses) compared to R-FOREWARNS. This pattern is also repeated for each of the severity level warning areas (minimal, minor and significant). However, during the testbed, participants only reported N-FOREWARNS's severity level warnings to be too high 8% of the time. The discrepancies between participant responses and the quantitative evaluation against R-FOREWARNS may reflect the error in the evaluation of participants, when verifying the SWF nowcasts. If so, this has important implications for the way surface water flood warnings are communicated - it is important to ensure that information is not misunderstood. Alternatively, the discrepancies may be due to the participants using the other available observation sets (Figure 6.7) to evaluate the performance of N-FOREWARNS, which led them to their responses (as opposed to error in judgement). It is important to note that the same model is used for R-FOREWARNS and N-FOREWARNS, meaning this evaluation is a fundamentally a comparison of their rainfall inputs.

The requirement to align the SWF nowcasts with the corresponding RFG warnings (see section 6.2.4) meant that, for 14 of the 20 testbed events, N-FOREWARNS (and SWFHIM) were assessed with lead times of 3 – 4 hours. However, STEPS skill has been shown to reduce rapidly over the first 3 hours, after which there is a transition from radar to NWP and skill flattens with lead time (Bowler et al. 2006). Table 6.3 shows the N-FOREWARNS output for the same five events in Table 6.2 but using a 1 hour lead time nowcast instead. For these cases, by driving it with a shorter lead time rainfall nowcast from STEPS (i.e. one with more skill), N-FOREWARNS is better able to capture the flood impact risk. Therefore, in order to have a better understanding of the performance of N-FOREWARNS, it is necessary to evaluate it across multiple lead times.

Table 6.3: Comparison of N-FOREWARNS maps issued during the testbed (with 2-3 hour lead times, see Table 6.2) against N-FOREWARNS with a 1 hour lead time and R-FOREWARNS, for each event (rows).

Figure 6.12a shows the distribution of POD, 1-POFD and SR spatial scores for N-FOREWARNS over 1 – 6 hours' lead time, for the 70 most significant SWF events recorded in Northern England from social media reports between 2013 - 2022 (taken from Maybee et al. 2024), verified against

R-FOREWARNS. Higher values denote higher skill for all three scores in Figure 6.12a (hence the use of 1–POFD). The likelihood of detecting a radar-identified event (i.e. mean POD score) decreases from 0.37 to 0.1 across 1 – 3 hours (as the skill of the STEPS nowcast reduces). The majority of scores then drop to 0 from 4 hours onwards. Across all lead times, the likelihood of falsely predicting an event is low (1–POFD score remains above 0.95), however, this result is to be expected due to the low POD scores (i.e. N-FOREWARNS produces a high proportion of no-SWF event predictions). When N-FOREWARNS produces a prediction, the likelihood that it is a hit (i.e. mean SR score) steadily reduces from 0.41 to 0.05 across 1 – 5 hours. The 0.05 increase in the mean SR score from 5 – 6 hours lead time may be a reflection of the increasing skill in the NWP model as it spins up (Clark et al. 2016).

To gain insight into N-FOREWARNS's performance over a full summer season, rather than evaluating it solely at the time of known flood events as in Figure 6.12a, N-FOREWARNS was run every hour between 1st May 2024 and 30th September 2024 and verified against R-FOREWARNS. Figure 6.12b shows the distribution of the skill scores for 3,740 issuances for lead times of 1 to 6 hours. Overall, the results are similar to those for the 70 known flood events in Figure 6.12a. The mean POD score decreases with lead time from 0.35 to 0.1, whilst the mean 1–POFD scores remain 1 across all lead times (93% of R-FOREWARNS issuances contained no events). The SR score decreases with lead time from 0.36 to 0.05 but, due to a slightly greater number of false alarms, the mean SR values are lower at each lead time compared to Figure 6.12a (by a difference of 0.05 for 1 hours' lead time). Testing on a dataset with a lower proportion of observed significant events means that false alarms are more likely to occur (compared to a dataset containing significant events only). Overall, the comparable results of Figure 6.12a and b shows that, even when it applied to a majority of cases with no SWF events, N-FOREWARNS still maintains its level of capability in capturing the rarer SWF events and does not have a high probability of false detection.

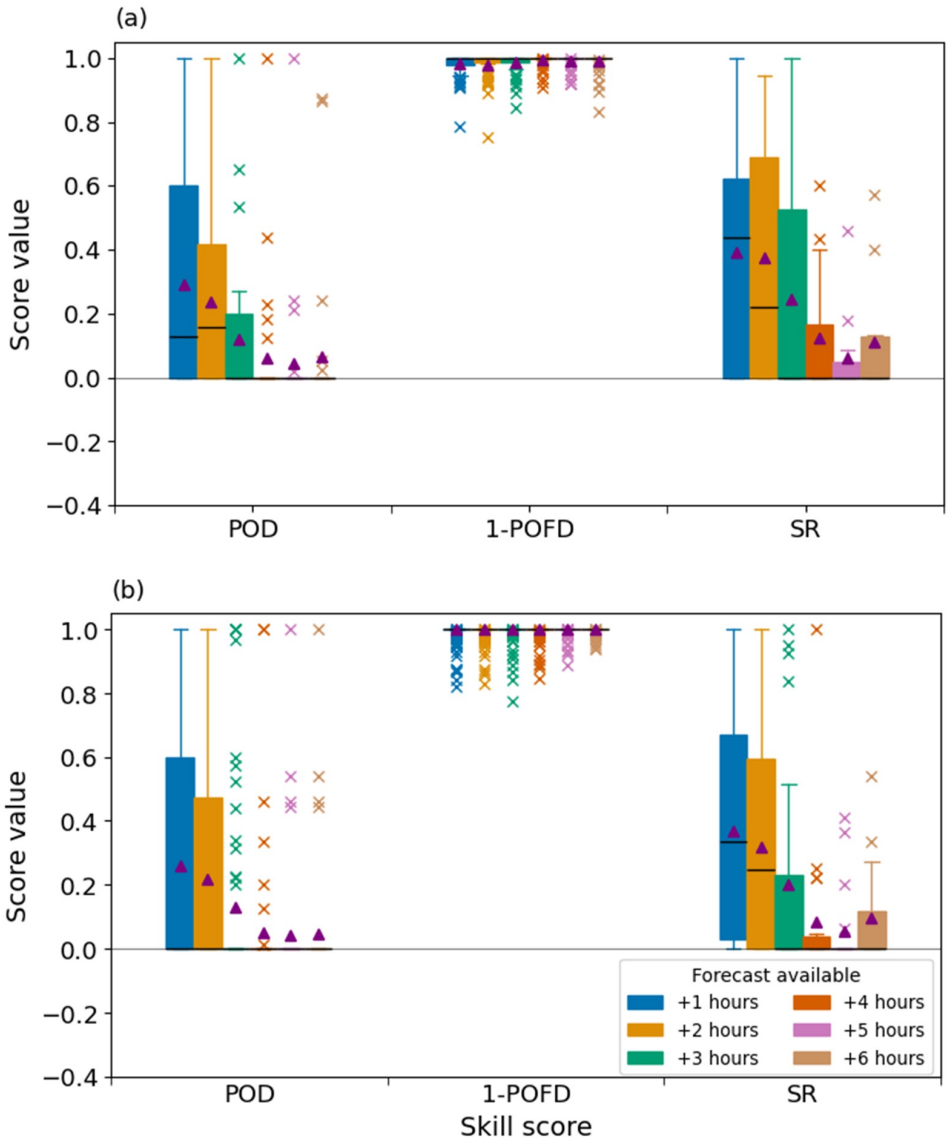


Figure 6.12: POD, 1-POFD and SR score distributions for a) 70 flooding events between 2013 – 2022 (to match Maybee et al., 2024) and b) 3,470 issuances across the continuous time period of summer 2024, produced by N-FOREWARNS for 1 – 6 hour lead times. Mean scores are represented by purple triangles, whilst the median scores are represented by horizontal black lines.

6.4 Conclusions and discussion

Access to accurate SWF predictions at short lead times is a key requirement in the flood response community. This study provides an evaluation of tools used for SWF prediction on nowcasting timescales. In particular, this study presents N-FOREWARNS, a new tool that produces SWF nowcasts at lead times of 1 to 6 hours using probabilistic rainfall predictions from the STEPS nowcast system. The results from this study provide both a qualitative evaluation

of N-FOREWARNS and SWFHIM (another SWF prediction tool) by flood professionals during the summer 2024 testbed, as well as a quantitative evaluation of N-FOREWARNS over a longer historical period.

A testbed hosted by the FFC and Met Office during summer 2024 focussed on nowcasting SWF capabilities with the aim to 1) compare the performance of N-FOREWARNS with SWFHIM (FFC's current SWF model) and 2) assess whether N-FOREWARNS and SWFHIM could add value to the Rapid Flood Guidance (RFG) production process, which was being trialled operationally over summer 2024. Across 20 SWF events during summer 2024, N-FOREWARNS produced fewer flood predictions compared to SWFHIM and missed a larger proportion of the observed SWF events, particularly at the minor impact severity level. However, when N-FOREWARNS did produce a SWF prediction it was generally accurate, with a low number of false alarms reported. By producing more predictions, SWFHIM achieved a greater number of hits than N-FOREWARNS. However, output from SWFHIM often has low spatial precision. This results in the SWF nowcast producing predictions in the incorrect location, triggering a larger number of false alarms. Furthermore, SWFHIM produces impact severity levels that are too high compared to the reported SWF observations that were available. The testbed also showed that the RFG is generally good at identifying the region of SWF impacts, and the accompanying detail on rainfall amounts and impact likelihood provides useful information for end users. However, the RFG warnings tend to be too widespread and overestimate the level of impact severity. The conclusion for aim 2), therefore, is that both N-FOREWARNS and SWFHIM can add value to the SWF forecasting process, by providing operational forecasters with improved spatial accuracy and help distinguishing between impact severity levels.

Testbed participant responses on location and impact severity levels of SWFHIM and N-FOREWARNS produced a wide distribution of performances. One reason is that each tool showed better/worse performance under certain conditions i.e. N-FOREWARNS performed better during high rainfall accumulation events and SWFHIM performed better during minor impact rainfall events. Another reason is the differing opinions on what participants consider to be important characteristics of a SWF prediction. Some participants considered a high number of false alarms to be an acceptable compromise in order to get a hit, whereas others prioritised reducing the false alarms to achieve better reliability. These opposing opinions highlight the challenge of definitively assessing SWF nowcasts and demonstrates the importance of involving

experts from different technical backgrounds to provide a well-rounded evaluation.

In order to make a direct comparison with the RFG during the testbed, N-FOREWARNS was qualitatively evaluated for a single lead time between 1 – 4 hours only. Therefore, a quantitative analysis across all lead times between 1 and 6 hours was also performed to provide a more complete picture. When evaluated for 70 historical, significant, observed flood events, N-FOREWARNS's ability to capture SWF events gradually decreased between a 1 and 3 hour lead time, after which there was a significant drop in performance. A 3-hour lead time limit on the usefulness of N-FOREWARNS is therefore implied. The significant drop off is due to the transition from radar-extrapolated rainfall data to NWP rainfall data within the STEPS algorithm. At these short lead times the NWP input is still undergoing spin-up, which is known to result in a drop in skill (Clark et al. 2016). This demonstrates the sensitivity of N-FOREWARNS to its input – a skilful N-FOREWARNS prediction is reliant on accurate rainfall data.

When evaluated over a continuous period of summer 2024 (3,740 issuances for each lead time), N-FOREWARNS produced a similar set of results, with marginally higher (lower) H (SR) scores (due to a small increase in the number of false alarms). This performance consistency over every day in a summer season, including days when no flood events occurred, shows that N-FOREWARNS can capture the less frequent SWF events and maintain a low probability of false detection.

N-FOREWARNS is a modified version of FOREWARNS (Maybee et al. 2024), which uses rainfall inputs from ensemble NWP forecasts and produces predictions at lead times of 1 to 4 days. A direct comparison of the skill in FOREWARNS and N-FOREWARNS skill is not possible, but it is still useful to qualitatively do so for context. For the 70 recorded flood events used in this study, N-FOREWARNS (at a 1 hour lead time) produces lower spatial POD scores (by 0.1) but higher 1-POFD and SR scores (by 0.20 and 0.10, respectively) than the FOREWARNS predictions at a 1 day lead time (evaluated in Maybee et al. (2024), who refer to POD as H and POFD as F). This suggests that N-FOREWARNS produces a lower proportion of false alarms, but at the compromise of fewer hits. When comparing across a continuous set of events (Maybee et al. 2024 evaluated over 725 continuous days for May – October 2019 – 2022), N-FOREWARNS produces approximately the same POD score but higher SR scores (by 0.20), which again can be contributed to a lower number of false alarms. Given that N-FOREWARNS is predicting

with a lower lead time than FOREWARNS (hours compared to days), it may be expected to produce better performance scores. However, the two methods have been set different predictive expectations – N-FOREWARNS is required to predict at a specific valid time (meaning that it needs to get the hour of the day correct), whereas FOREWARNS is predicting events that occur within a 24-hour period. Having to overcome these inherently different challenges has resulted in comparable scores for the two SWF prediction tools. Nevertheless, both tools show less than 50% chance of capturing their corresponding radar-driven SWF risks (POD scores), highlighting that, despite the short lead times of N-FOREWARNS, skill scores for SWF prediction are still low. Significant improvements in the ability of NWP and nowcasting to predict the location of convective rainfall is required to improve SWF prediction.

Through expert evaluation and quantitative verification, this study has shown that N-FOREWARNS would be a valuable tool that summarises SWF risk for any given hour in a single map for an operational hydrometeorologist producing SWF predictions. Overall, N-FOREWARNS's skill extends up to a 3 hour lead time and its performance is highly dependent on the accuracy of the nowcast rainfall data.

Chapter 7

Conclusions

Organised convective storms generate hazardous weather conditions that have both societal and economic impacts. Effective early warning weather systems require the capability to make short-term predictions (minutes to hours ahead) of convective storms – a major challenge in the field of meteorology – which NWP alone struggles to fulfil. Nowcasting rapidly generates short-term predictions of the future atmosphere to fill the NWP capability gap. Recent nowcasting advancements have trained machine learning models on weather radar data to predict future radar frames for capturing storm development. Although these approaches are continuing to increase in skill, their capability in capturing isolated processes such as convective initiation, has not been demonstrated. Furthermore, the sparsity of weather radar data across the Tropics is a barrier to accessing established nowcasting techniques.

To effectively communicate a short-term weather warning it must be clear and easily interpretable. The translation of crude nowcasting information e.g. predicted rainfall, into a more relatable quantity e.g. likelihood of flooding, enables forecasters and emergency responders to make more informed decisions. This crucial part of end-to-end early warning provision is often overlooked in nowcasting research.

The purpose of this thesis is to address the aforementioned challenges in the nowcasting field. More specifically, the research questions that were posed in this thesis are:

1. Can geostationary satellite data be used for skilful optical flow nowcasting over the Maritime Continent?

2. Can satellite-based ML techniques be used to nowcast isolated convective processes in the

Maritime Continent, and elsewhere in the Tropics?

3. Can a surface water flood nowcasting tool provide added value over England and Wales?

7.1 Addressing the research questions

Chapter 4 focuses on evaluating the use of BT data from the Himawari 8/9 satellite for optical flow-based nowcasting over the MC. The LK (deterministic) and STEPS (probabilistic) optical flow algorithms are used to produce nowcasts for lead times of up to 6 hours. When evaluated on a 10 km scale, the LK algorithm produces skilful nowcasts for up to 4 hours' lead time. Skill maps show that the LK algorithm performs best over regions where convection tends to propagate e.g. over the sea, and poorly over regions where the initiation, growth and decay of convection is occurring. Although this is a well-known downfall of optical flow models (Burton et al. 2022; Hill et al. 2020), explicit spatial mapping of optical flow skill in this way – identifying timing and location of nowcast usefulness – has not been previously presented. The STEPS algorithm produces skilful and reliable ensemble nowcasts for 3 hours' lead time, but with an under-dispersive distribution (i.e. low ensemble member variance). In line with the LK algorithm, STEPS produces its lowest skill over the land during the afternoon and highest skill over the sea during the morning. An overall comparison of STEPS, LK and a persistence baseline shows that, when optimally thresholded, STEPS performs the best across 1 – 6 hours' lead time (on a 10 km evaluation scale).

To address some of the downfalls of optical flow nowcasting that were identified in chapter 4, chapter 5 presents Simple Initiation and Intensification Nowcasting neural Network (SII-NowNet): a machine learning tool for nowcasting convection initiation and intensification in the Tropics. Over Sumatra, Indonesia, SII-NowNet produces useful skill for up to 3 hours' lead time for intensification and 2 hours' lead time for initiation. Beyond these limits, it begins to over-predict the convective events (i.e. it produces false alarms). To exemplify applicability in other Tropical regions, SII-NowNet is tested over Zambia where, without retraining or region-specific tuning, SII-NowNet achieves skill scores comparable to those over Sumatra. Sensitivity testing (over Sumatra) shows that the number of training samples can be reduced to three weeks for intensification (99% reduction in training time) and three months for initiation (75% reduction in training time), before its outperformed by climatology. Furthermore, testing the relative importance of SII-NowNet's input channels suggests that satellite BT data from the

prior hour can still contribute useful information for predicting convection up to 3 hours ahead.

Finally, chapter 6 moves away from Tropical nowcasting to focus on developing a surface water flood nowcasting tool for England and Wales. N-FOREWARNS (a development of the FOREWARNS tool (Maybee et al. 2024)) uses STEPS ensemble rainfall nowcasts to produce hourly SWF nowcasts for lead times of 1 – 6 hours. Expert user feedback from a nowcast comparison testbed (which included current operational SWF prediction tools) indicates that N-FOREWARNS produces few false SWF warnings. Furthermore, when warnings are issued by N-FOREWARNS, they have high spatial accuracy, with users noting an incorrect warning location only 7% of the time. However, users also report that N-FOREWARNS misses a high proportion of SWF events with minor impacts, caused by smaller-scale convective events. Overall, the testbed shows that N-FOREWARNS can add value to the SWF forecasting process, by providing operational forecasters with improved spatial accuracy and help distinguishing between SWF severity levels. Quantitative verification shows that N-FOREWARNS's 1 hour lead time nowcasts produce useful skill, but N-FOREWARNS reaches a limit of predictability at 3 hours' lead time, and that the skill is primarily limited by the accuracy of its rainfall inputs.

7.2 Broader implications

The performance of radar-based optical flow algorithms for nowcasting is well understood, particularly in the mid-latitudes. Instead, this thesis shows that, over the MC, satellite BT data can be used as supplementary input to optical flow algorithms for continuous nowcasting. This capability can provide benefits across the Tropics, where radar data is unavailable/unreliable due to economic restraints. It is well known that optical flow algorithms are unable to capture developing convection; however, the explicit mapping of the timing and location of optical flow skill over the MC provides a useful tool for forecasters, informing them when and where these nowcasting methods are suitable for operational use.

To the author's knowledge there are currently no other ML tools that produce probabilistic nowcasts (on 1 – 6 hour lead times) of convection initiation and convection intensification using solely satellite BT data. The performance capability of SII-NowNet has shown that ML techniques can provide predictability for these isolated convective events, filling some of the previously identified skill gaps in optical flow nowcasting. The use of freely available satellite inputs, combined with SII-NowNet's ease of applicability to other Tropical regions, makes it a suitable tool for op-

erationalisation. Currently, Badan Meteorologi, Klimatologi, dan Geofisika (BMKG; Indonesian national weather service) have adopted SII-NowNet and are running it in an operational trial, which can be seen on their website (https://web-meteo.bmkg.go.id/id/peringatan/sii-nownet). The South African Weather Service (SAWS) will also begin implementation of SII-NowNet into their nowcasting operations during summer 2025 (see section 7.3 for further details).

Many national weather services in the Tropics have comparatively little funding (but experience some of the most impactful weather) compared to the mid-latitudes, meaning that the storage of large datasets is not always possible. SII-NowNet's ability to perform well in new locations without any re-training removes the requirement for large computer facilities. Furthermore, if re-training is required, SII-NowNet's ability to maintain acceptable performance when trained on a reduced dataset size makes it an accessible tool for operationalisation in resource-constrained settings.

The comparable performance of SII-NowNet over a new region with different meteorological conditions also provides insight into our understanding of convection. Controls on convection differ greatly between Sumatra and Zambia, which SII-NowNet is given no knowledge of. However, through the encoding/decoding stages, the model is able to identify patterns and features of convection on a localised scale (using a 3 x 3 pixel filter) over both regions. This suggests that, although larger-scale forcings and triggering mechanisms differ between Sumatra and Zambia, the localised patterns of newly initiating and intensifying convection show similarities.

With previously little focus given to nowcasting hazardous events due to rainfall, N-FOREWARNS provides a key step towards improving capability for predicting SWF risk for 1 – 6 hours ahead. By producing single SWF risk maps (as opposed to multiple impact maps), N-FOREWARNS provides clear and concise information that is accessible for decision makers. Hourly predictions with hourly lead times enables live and up-to-date SWF risk information, which is crucial for emergency services' preparedness.

The testbed results also highlight the importance of gathering expert user feedback when evaluating prediction tools. Differing opinions on 'what makes a good SWF forecast?' provided a well-rounded evaluation of the performance of the SWF tools. This method of evaluation is recommended for future testing of nowcasting tools. Although N-FOREWARNS was received well during the testbed, quantitative testing clearly shows that there is still a performance limit in capturing SWF event risk. Continued focus on improving rainfall nowcasting is required to

overcome this limit.

7.3 Limitations and future work

Geostationary satellites provide a means for accessible nowcasting in the Tropics; however, they are constrained to a single view of the Earth from above, limiting the amount of information that can be inferred from their images. From cloud top BT images, the distribution of the coldest clouds (assumed to be the deepest in the troposphere) can be clearly identified, but their characteristics (e.g. stratiform or convective rainfall) are ambiguous. Previous literature has used pre-defined BT thresholds to isolate the coldest clouds that are associated with large scale ascent (i.e. a proxy for convection) (Crook et al. 2024; Feng et al. 2021; Machado and Laurent 2004; Roca et al. 2017). However, non-convective clouds that are also colder than this threshold e.g. cirrus clouds, will be included in this proxy. Therefore, the use of inaccurate proxies for convection provides one of the main sources of limitation in this thesis. Chapter 4 applies a 235 K threshold to produce optical flow nowcasts of estimated convection (that likely include both convective and non-convective clouds), whilst chapter 5 uses a range of BT thresholds for developing the initiation and intensification identification methodology (with limitations highlighted by the highly sensitive threshold for initiation identification).

To address these limitations, future work should focus on improving the methods to isolate convection. One solution would be to deploy the wavelet decomposition method developed by (Klein et al., 2018), which has shown skill in isolating convective regions of storms in satellite BT data over West Africa. For the initiation identification method in chapter 5, a reverse approach could be taken, whereby a storm is identified at its mature stage and then tracked backward in time to its initiation – this has previously been attempted with radar data over Korea by (Han et al., 2019). Alternatively, precipitation data from the Global Precipitation Measurement (GPM) mission could replace the BT data used in this thesis, enabling direct rainfall nowcasting (BT cannot provide direct information on rainfall). Integrated Multi-satellitE Retrievals for GPM (IMERG) (Hou et al. 2014) is a dataset of global precipitation estimates that uses microwave observations, infrared estimates, and surface data. However, the IMERG dataset does not provide real-time data, meaning that it is not an operationally viable option for chapter 4. However, it could be used in the identification methods of chapter 5 to build a dataset of initiating and intensifying rainfall events. An important acknowledgement to make

is that IMERG data has lower resolution (10 km) than the BT data in this study (i.e. it would decrease the resolution of SII-NowNet), and has known biases such as overestimation of heavy rainfall events (Zhang et al. 2023b). Rain gauges, however, do provide an accurate source for rainfall observations. Although sparsely distributed across the Tropics, their measurements can be used for verification and calibration at point locations (Taromideh et al. 2025).

Chapter 4 identified the initiation-growth-decay problem within optical flow nowcasting. SII-NowNet aims to solve this for initiation and growth but does not address decay – key information for decision-making e.g. saving emergency services resources. A direction of future work could be to use BT warming as a proxy to investigate SII-NowNet's capability in predicting decaying events. Another direction is to focus on testing SII-NowNet's inputs. The use of only two BT images in chapter 5 enabled computationally efficient model training and data processing. However, adding greater temporal frequency to the inputs (e.g. BT images every 15 minutes) may add more information and improve skill. The addition of BT images from before the previous hour may also provide useful information for predicting future convection development. During daylight hours visible imagery is an available observation that could be used to highlight convergence lines and uplift (Birch et al. 2014). The Global Forecasting System (National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce 2015) produces live NWP output, which provides a range of atmospheric variables that could be incorporated and tested as SII-NowNet inputs, such as wind shear and vertical velocity. Alternatively, combining optical flow fields, initiation/intensification observations and BT images may provide rich input information for nowcasting the timing and propagation of developing convection.

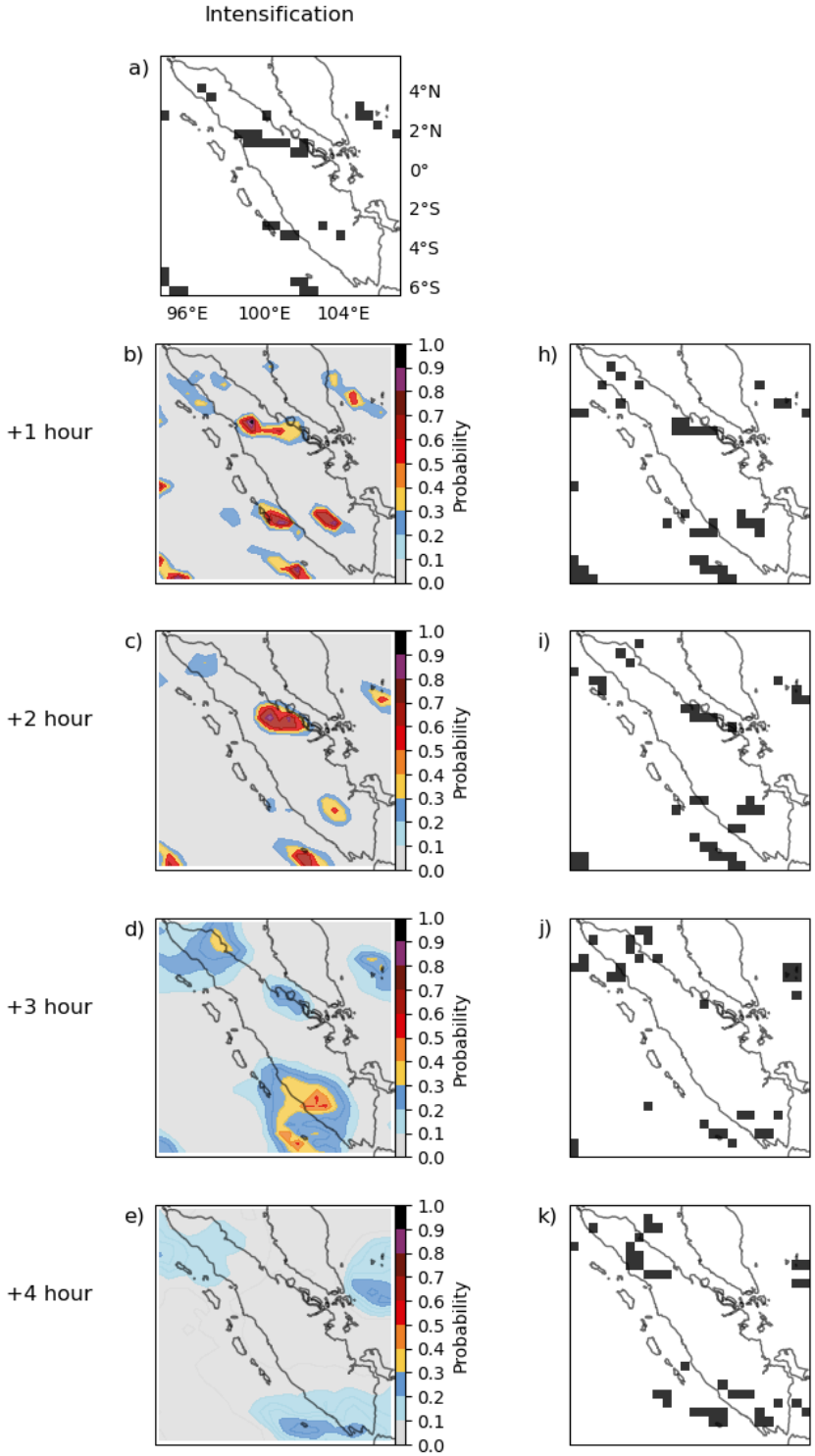
The immediate plans for SII-NowNet are to continue its implementation with national weather services in the Tropics – SII-NowNet is freely available for anyone to use. The next step in the current collaboration with BMKG is to increase its spatial coverage across the entire MC. This larger domain will require re-training with an increase in GPU compute, but by applying transfer learning the training dataset size can be reduced (compared to the original training dataset). When applied to Zambia, SII-NowNet showed that it could provide comparable skill. However, no tuning was made to the identification methodologies over this region to account for the larger BT field grid spacing. Further work over Africa should focus on recalibrating the identification methodologies parameters for more region-specific optimization. A collaborative project with

SAWS (the Regional Specialised Meteorological Centre for Southern Africa) in summer 2025 aims to implement SII-NowNet over South Africa and across southern Africa. With some availability of radar coverage across South Africa, more detailed testing and development of SII-NowNet will be possible. Furthermore, as part of the SAWS project, a period of ML-based training will be provided to SAWS, transferring the skills and enabling them to take ownership of SII-NowNet for further development if desired.

In chapter 6, N-FOREWARNS predicts flood return periods, to produce a single, simple SWF risk map; however, this simplicity may not fulfil some user-needs that require specific details of impact severity. Qualitative testing with decision-makers too (rather than just technical experts) would provide more clarification on the key priorities. Furthermore, the quantitative analysis performed in chapter 6 provides an understanding of how N-FOREWARNS performs against a corresponding radar-driven version that assumes a 'perfect' rainfall field. It is therefore an assessment of the rainfall fields, not the SWF impacts themselves. Ideally, N-FOREWARNS output would be verified against a reliable dataset of SWF observations. However, deficiencies in collating SWF observations (of both the impacts and the floods themselves) (e.g. unreliable sources) mean that a robust evaluation is undermined. To overcome this, future work should focus on the improvement of SWF impact identification methodologies. The use of satellite data has shown promise for this application (Notti et al. 2018), specifically combined with machine learning tools for flood mapping (Portalés-Julià et al. 2023).

Finally, a collaboration over summer 2025 with the Met Office will implement N-FOREWARNS into their operations. This will allow opporunity to test the sensitivity of N-FOREWARNS's paramters. For example, decreasing the 98th rainfall percentile parameter will allow more rainfall into the nowcast, which may help to reduce the number of N-FOREWARNS misses. Furthermore, through this project there is potential to use more skilful rainfall nowcasts as input products (e.g. MONOW; Moseley and Sandford 2019), which would address the key limitation of N-FOREWARNS.

Appendix A


Intensification
a)
4°N
2°N
0°
2°S
4°S
6°S
96°E
100°E
104°E
b)
h)
+1 hour
c)
i)
+2 hour
d)
j)
+3 hour
e)
k)
+4 hour
Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

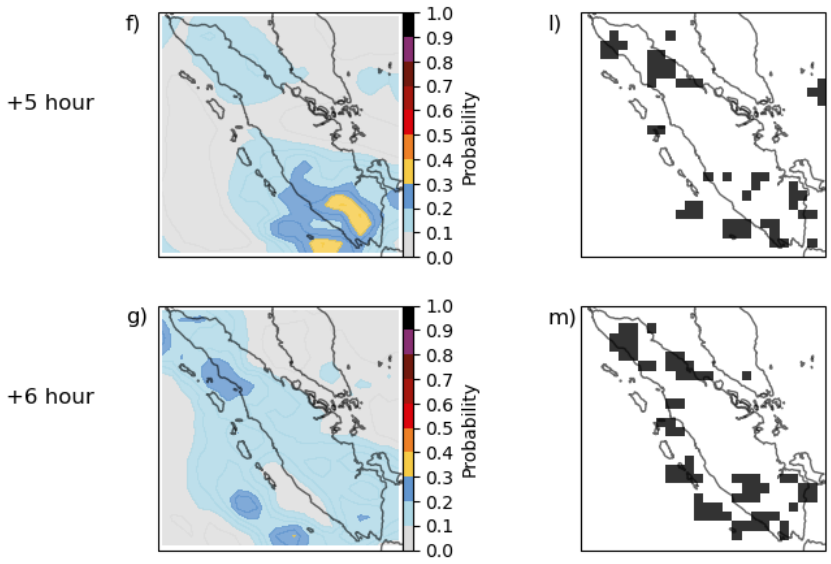
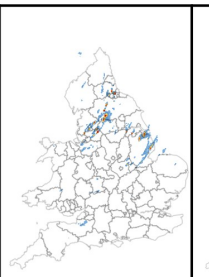
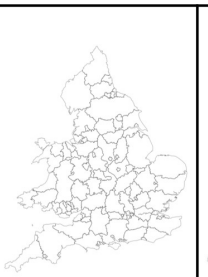
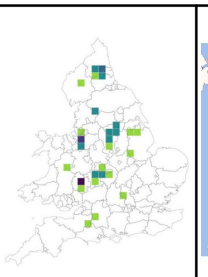
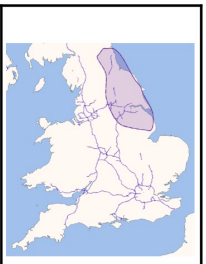
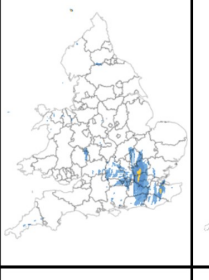
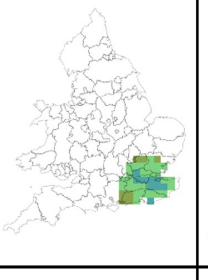
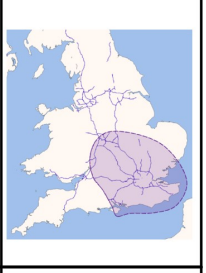
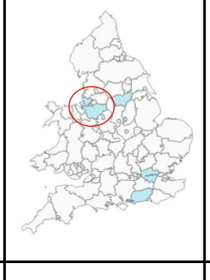
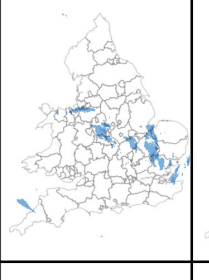
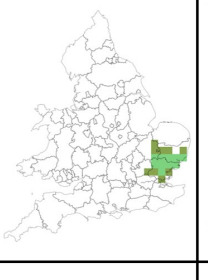
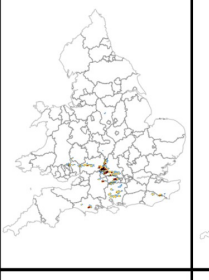
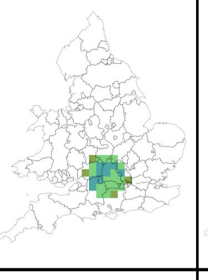
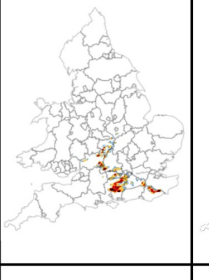
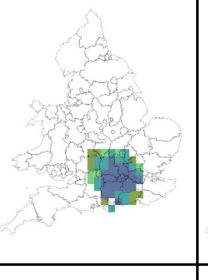
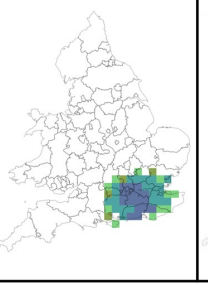
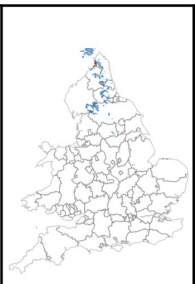
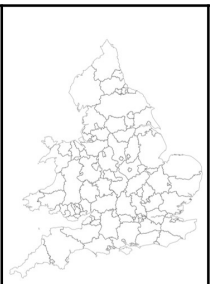
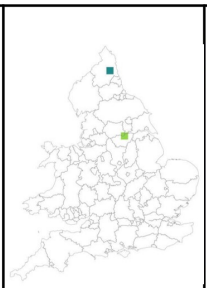
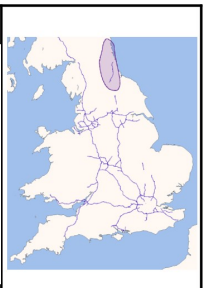
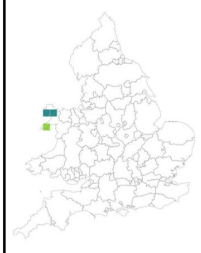
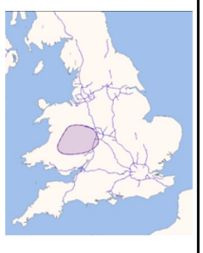
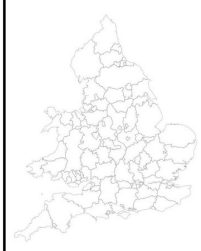
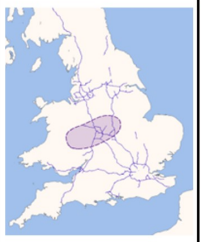
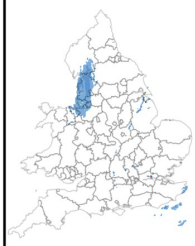
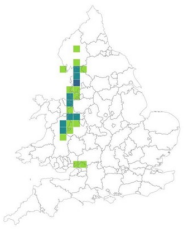
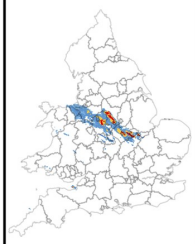
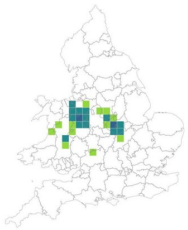
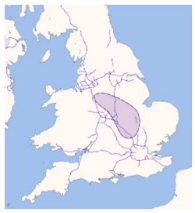
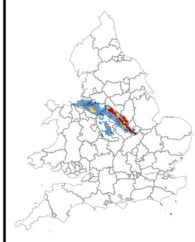
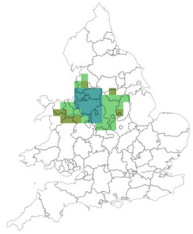
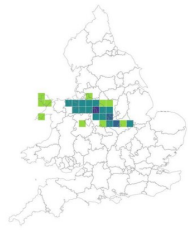
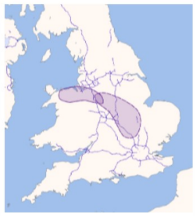
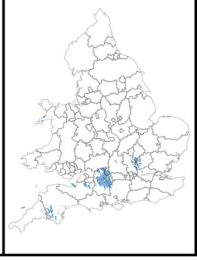
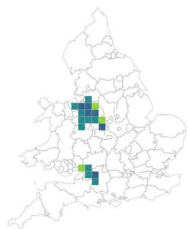
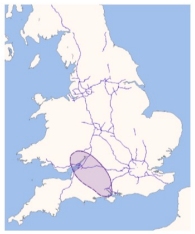


Figure A.1: Same case study as Figure 5.8 but for SII-NowNet's intensification nowcasts b) - e) compared against the corresponding persistence nowcasts (h - k).

	Observed Impact Minimal Minor Significant Severe	3 hour radar accumulations (>10mm) 10 20 30 40 50 60 70 80 90 100 3 hour rainfall accumulation/mm	N-FOREWARNS Return period/years 5 (= Minimal) 10 (= Minor) 30 (= Minor) 100 (= Significant) 1000 (= Severe)	SWFHIM (minor) 1 member Very low (2-5 members) Low (6-10 members) Medium (11-15 members) High (16+ members)	RFG (warning area enclosed by polygon)
Area 1 15th Jun 2024 Initialised: 1300 UTC Valid until: 1600 UTC (+3 hour lead time)					
Area 2 18th Jun 2024 Initialised: 1500 UTC Valid until: 1800 UTC (+3 hour lead time)					
Area 3 12th May 2024 Initialised: 1500 UTC Valid until: 1800 UTC (+3 hour lead time)					

Area 4 26th May 2024 Initialised: 1500 UTC Valid until: 1800 UTC (+3 hour lead time)					
Area 5 15th Jul 2024 Initialised: 1800 UTC Valid until: 2100 UTC (+3 hour lead time)					
Area 6 15th Jul 2024 Initialised: 2100 UTC Valid until: 2400 UTC (+3 hour lead time)					
Area 7a 1st Aug 2024 Initialised: 1500 UTC Valid until: 1600 UTC (+1 hour lead time)					
Area 7b 1st Aug 2024 Initialised: 1600 UTC Valid until: 1800 UTC (+2 hour lead time)					
Area 7c 1st Aug 2024 Initialised: 1900 UTC Valid until: 2000 UTC (+1 hour lead time)					

Area 8 16th Jul 2024 Initialised: 1500 UTC Valid until: 1700 UTC (+2 hour lead time)					
Area 9a 2nd Sep 2024 Initialised: 1400 UTC Valid until: 1700Z (+3 hour lead time)					
Area 9b 2nd Sep 2024 Initialised: 1700 UTC Valid until: 2000 UTC (+3 hour lead time)					
Area 10a 8th Sep 2024 Initialised: 0900 UTC Valid until: 1100 UTC (+2 hour lead time)					
Area 10b 8th Sep 2024 Initialised: 1000 UTC Valid until: 1300 UTC (+3 hour lead time)					

Area 11a 8th Sep 2024 Initialised: 1400 UTC Valid until: 1600 UTC (+2 hour lead time)					
Area 11b 8th Sep 2024 Initialised: 1600 UTC Valid until: 1900 UTC (+3 hour lead time)					
Area 12a 21st Sep 2024 Initialised: 1500 UTC Valid until: 1800 UTC (+3 hour lead time)					
Area 12b 21st Sep 2024 Initialised: 1700 UTC Valid until: 2000 UTC (+3 hour lead time)					
Area 13a 22nd Sep 2024 Initialised: 1200 UTC Valid until: 1500 UTC (+3 hour lead time)					

Area 13b 22nd Sep 2024 Initialised: 1600 UTC Valid until: 1900 UTC (+3 hour lead time)	

Table A.1: Same as Table 1 but for all areas evaluated across the testbed. The term 'areas' is used instead of 'Event' here because some events were evaluated multiple times during their validity period.

Q#	Question
1	What is the highest impact level of your N-FOREWARNS area?
2	What is the highest impact level of your SWFHIM area?
3	What (if any) was the main difference between your nowcast polygons based on SWFHIM and N-FOREWARNS? • Impact severity • Polygon size • Polygon location • Little difference • Only drew polygon for one model
4	Overall, how would you assess N-FOREWARNS and SWFHIM model output (not polygons)? • Hits (events that are forecast and observed) • Misses (observed events which were not forecast) • False Alarms (events that were forecast but not observed) • Correct Rejections (not forecast and not observed)

5	How well does the assessment you gave in the previous question represent the forecast? • Not very representative • Somewhat representative • Very representative
6	Did your N-FOREWARNS/SWFHIM/RFG polygon identify the right severity of impacts? • False alarm (0) • Much too high (1) • Slightly too high (2) • About right (3) • Slightly too low (2) • Much too low (1) • Miss (0)
7	Did your N-FOREWARNS/SWFHIM/RFG polygon identify the right location of impacts? • Incorrect location (0) • Much too large (1) • Slightly too large (2) • About right (3) • Slightly too small (2) • Much too small (1) • Miss (0)
8	What are the main differences between your N-FOREWARNS polygons and the RFG polygon? (select all that apply) • Impact Severity • Location • Coverage • Little difference • Did not draw N-FOREWARNS polygon

9	What are the main differences between your SWFHIM polygon and the RFG polygon? (select all that apply) • Impact Severity • Location • Coverage • Little difference • Did not draw N-FOREWARNS polygon
10	Would access to N-FOREWARNS or SWFHIM model output have improved the RFG polygon? • Yes • No • In some areas • Unsure
11	Please provide your overall subjective assessment of N-FOREWARNS performance for this event/area (1 = very poor, 5 = very good)
12	Please provide your overall subjective assessment of SWFHIM performance for this event/area (1 = very poor, 5 = very good)
13	Please provide your overall subjective assessment of RFG performance for this event/area (1 = very poor, 5 = very good)

Table A.2: The survey questions that were given to participant groups during the testbed.

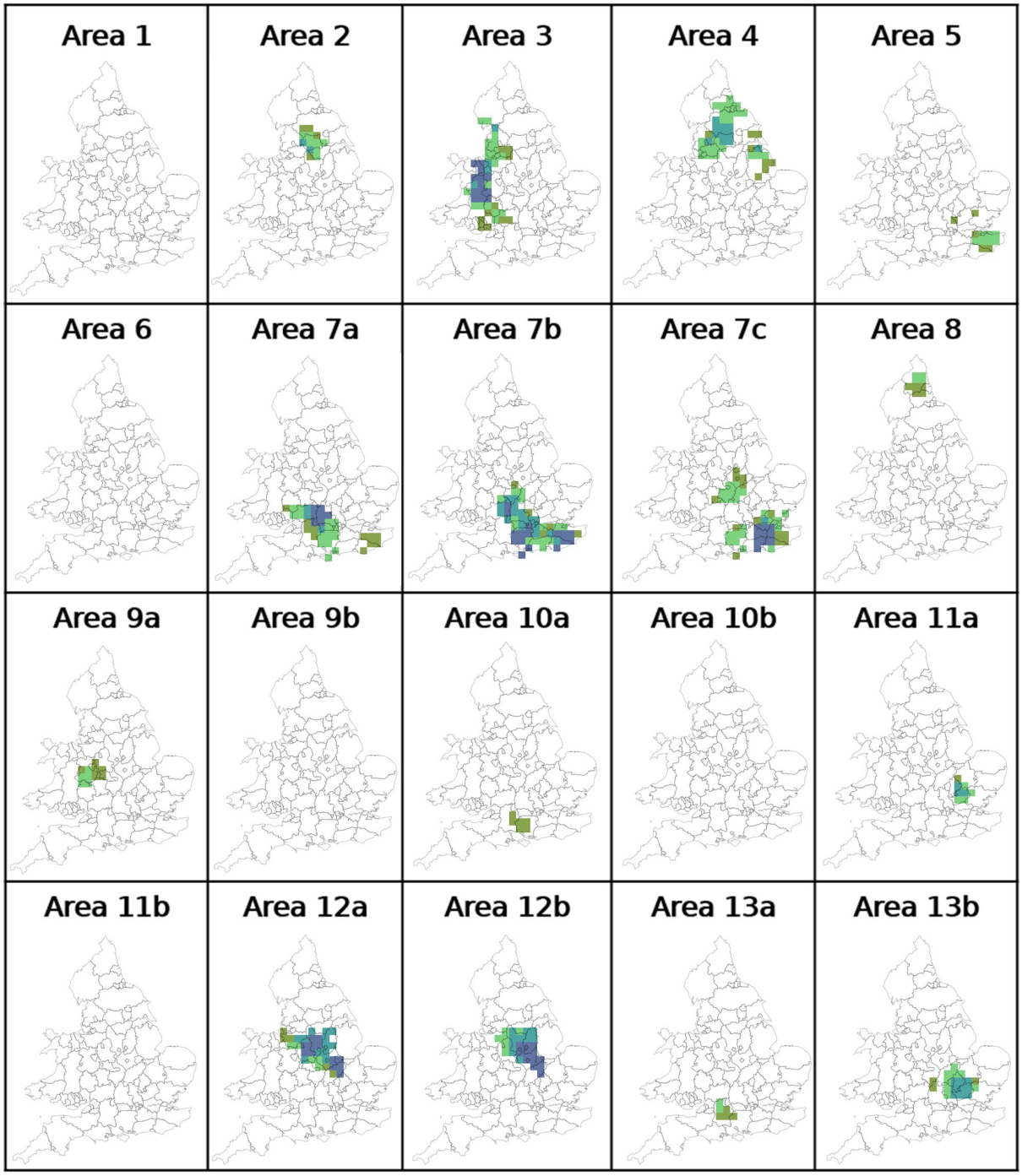


Table A.3: R-FOREWARNS output for each area evaluated in the testbed. Each 'Area' corresponds with the 'Area' in Table A.1.

Bibliography

Acosta-Coll, M., Ballester-Merelo, F., Martinez-Peiró, M., and De La Hoz-Franco, E. (2018). Real-Time Early Warning System Design for Pluvial Flash Floods—A Review. *Sensors*, 18(7):2255.

Adler, R. F., Gu, G., Sapiano, M., Wang, J.-J., and Huffman, G. J. (2017). Global Precipitation: Means, Variations and Trends During the Satellite Era (1979–2014). *Surveys in Geophysics*, 38(4):679–699.

Agarap, A. F. (2019). Deep Learning using Rectified Linear Units (ReLU). arXiv:1803.08375 [cs].

Ageet, S., Parker, D., Msemo, H. E., Colenbrander, W., de Coning, E., Gijben, M., Gonçalves, J., Groves, D., Imbwae, F., Juhanko, K., Kieth, G., Kroese, N., Marsham, J., Mbuli, N., Gilda, M., Monjane, M., Murambadoro, M., Perrels, A., Roberts, A., Sebego, E., Segalwe, O., Stander, J., and Tawane, T. (2024). Weather and Climate Information Services Early Warning for Southern Africa (WISER-EWSA) testbed 1. Technical report, University of Leeds.

Agrawal, S., Barrington, L., Bromberg, C., Burge, J., Gazen, C., and Hickey, J. (2019). Machine Learning for Precipitation Nowcasting from Radar Images. *arXiv:1912.12132 [cs, stat]*. arXiv: 1912.12132.

Aldrian, E. and Dwi Susanto, R. (2003). Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature. *International Journal of Climatology*, 23(12):1435–1452.

Aldridge, T., Gunawan, O., Moore, R. J., Cole, S. J., Boyce, G., and Cowling, R. (2020).

Developing an impact library for forecasting surface water flood risk. *Journal of Flood Risk Management*, 13(3):e12641.

Ali, A., Supriatna, S., and Sa'adah, U. (2021). RADAR-BASED STOCHASTIC PRECIPITATION NOWCASTING USING THE SHORT-TERM ENSEMBLE PREDICTION SYSTEM (STEPS) (CASE STUDY: PANGKALAN BUN WEATHER RADAR). *International Journal of Remote Sensing and Earth Sciences (IJReSES)*, 18(1):91.

Amell, A., Hee, L., Pfreundschuh, S., and Eriksson, P. (2025). Probabilistic near real-time retrievals of Rain over Africa using deep learning.

Autones, F. (2016). NWC-CDOP2-GEO-MFT-SCI-ATBD-Convection_v1.1. (1):72.

Ayzel, G., Scheffer, T., and Heistermann, M. (2020). RainNet v1.0: a convolutional neural network for radar-based precipitation nowcasting. *Geoscientific Model Development*, 13(6):2631–2644.

Baker, L. (2017). A Summer of Floods !

Barrett, P., Abel, S., Lean, H., Price, J., Stein, T., Stirling, A., and Darlington, T. (2021). WesCon 2023: Wessex UK Summertime Convection Field Campaign.

Bedka, K. M. and Khlopenkov, K. (2016). A Probabilistic Multispectral Pattern Recognition Method for Detection of Overshooting Cloud Tops Using Passive Satellite Imager Observations. *Journal of Applied Meteorology and Climatology*, 55(9):1983–2005.

Bennett, L. J., Browning, K. A., Blyth, A. M., Parker, D. J., and Clark, P. A. (2006). A review of the initiation of precipitating convection in the United Kingdom. *Quarterly Journal of the Royal Meteorological Society*, 132(617):1001–1020.

Berg, P., Wagner, S., Kunstmann, H., and Schädler, G. (2013). High resolution regional climate model simulations for Germany: part I—validation. *Climate Dynamics*, 40(1-2):401–414.

Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., Kumagai, Y., Miyakawa, T., Murata, H., Ohno, T., Okuyama, A., Oyama, R., Sasaki, Y., Shimazu, Y., Shimoji, K., Sumida, Y., Suzuki, M., Taniguchi, H., Tsuchiyama, H., Uesawa, D., Yokota, H., and Yoshida, R. (2016). An Introduction to Himawari-8/9— Japan’s New-Generation Geostationary Meteorological Satellites. *Journal of the Meteorological Society of Japan. Ser. II*, 94(2):151–183.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619(7970):533–538.

Birch, C. E., Marsham, J. H., Parker, D. J., and Taylor, C. M. (2014). The scale dependence and structure of convergence fields preceding the initiation of deep convection: Convergence and convection initiation. *Geophysical Research Letters*, 41(13):4769–4776.

Birch, C. E., Webster, S., Peatman, S. C., Parker, D. J., Matthews, A. J., Li, Y., and Hassim, M. E. E. (2016). Scale Interactions between the MJO and the Western Maritime Continent. *Journal of Climate*, 29(7):2471–2492.

Bowler, N. E., Pierce, C. E., and Seed, A. W. (2006). STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Quarterly Journal of the Royal Meteorological Society*, 132(620):2127–2155.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Browning, K. A. (1980). Review Lecture: Local weather forecasting. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 371(1745):179–211.

Browning, K. A. and Collier, C. G. (1989). Nowcasting of precipitation systems. *Reviews of Geophysics*, 27(3):345–370.

Burton, R. R., Blyth, A. M., Cui, Z., Groves, J., Lamptey, B. L., Fletcher, J. K., Marsham, J., Parker, D. J., and Roberts, A. (2022). Satellite-based nowcasting of West African mesoscale storms has skill at up to four hours lead time. *Weather and Forecasting*.

Böing, S. J., Birch, C. E., Rabb, B. L., and Shelton, K. L. (2020). A percentile-based approach to rainfall scenario construction for surface-water flood forecasts. *Meteorological Applications*, 27(6):e1963.

Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P. W., Trisos, C., Romero, J., Aldunce, P., Barrett, K., Blanco, G., Cheung, W. W., Connors, S., Denton, F., Diongue-Niang, A., Dodman, D., Garschagen, M., Geden, O., Hayward, B., Jones, C., Jotzo, F., Krug, T., Lasco, R., Lee, Y.-Y., Masson-Delmotte, V., Meinshausen, M., Mintenbeck, K., Mokssit, A., Otto, F. E., Pathak, M., Pirani, A., Poloczanska, E., Pörtner, H.-O., Revi, A., Roberts, D. C., Roy, J., Ruane, A. C., Skea, J., Shukla, P. R., Slade, R., Slangen, A., Sokona, Y., Sörensson, A. A., Tignor, M., Van Vuuren, D., Wei, Y.-M., Winkler, H., Zhai, P., Zommers,

Z., Hourcade, J.-C., Johnson, F. X., Pachauri, S., Simpson, N. P., Singh, C., Thomas, A., Totin, E., Arias, P., Bustamante, M., Elgizouli, I., Flato, G., Howden, M., Méndez-Vallejo, C., Pereira, J. J., Pichs-Madruga, R., Rose, S. K., Saheb, Y., Sánchez Rodríguez, R., Ürge Vorsatz, D., Xiao, C., Yassaa, N., Alegría, A., Armour, K., Bednar-Friedl, B., Blok, K., Cissé, G., Dentener, F., Eriksen, S., Fischer, E., Garner, G., Guivarch, C., Haasnoot, M., Hansen, G., Hauser, M., Hawkins, E., Hermans, T., Kopp, R., Leprince-Ringuet, N., Lewis, J., Ley, D., Ludden, C., Niamir, L., Nicholls, Z., Some, S., Szopa, S., Trewin, B., Van Der Wijst, K.-I., Winter, G., Witting, M., Birt, A., Ha, M., Romero, J., Kim, J., Haites, E. F., Jung, Y., Stavins, R., Birt, A., Ha, M., Orendain, D. J. A., Ignon, L., Park, S., Park, Y., Reisinger, A., Cammaramo, D., Fischlin, A., Fuglestvedt, J. S., Hansen, G., Ludden, C., Masson-Delmotte, V., Matthews, J. R., Mintenbeck, K., Pirani, A., Poloczanska, E., Leprince-Ringuet, N., and Péan, C. (2023). IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland. Technical report, Intergovernmental Panel on Climate Change (IPCC). Edition: First.

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L., and Zdeborová, L. (2019). Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4):045002.

Carlyon, H., Hitching, J., McNeill, A., and Dixon, T. (2012). Northumberland County Council Sustainable Transport FCERM County Hall Morpeth NE61 2EF. Technical report, Northumberland County Council.

Chang, C.-P., Wang, Z., McBride, J., and Liu, C.-H. (2005). Annual Cycle of Southeast Asia—Maritime Continent Rainfall and the Asymmetric Monsoon Transition. *Journal of Climate*, 18(2):287–301.

Chih-Chung Chang, Chih-Wei Hsu, and Chih-Jen Lin (2000). The analysis of decomposition methods for support vector machines. *IEEE Transactions on Neural Networks*, 11(4):1003–1008.

Clark, P., Roberts, N., Lean, H., Ballard, S. P., and Charlton-Perez, C. (2016). Convection-permitting models: a step-change in rainfall forecasting: A step-change in rainfall forecasting.

Meteorological Applications, 23(2):165–181.

Clark, R. A., Gourley, J. J., Flamig, Z. L., Hong, Y., and Clark, E. (2014). CONUS-Wide Evaluation of National Weather Service Flash Flood Guidance Products. *Weather and Forecasting*, 29(2):377–392.

Cole, S. J. and Moore, R. J. (2009). Distributed hydrological modelling using weather radar in gauged and ungauged basins. *Advances in Water Resources*, 32(7):1107–1120.

Cole, T. J. (1991). Applied logistic regression. D. W. Hosmer and S. Lemeshow, Wiley, New York, 1989. *Statistics in Medicine*, 10(7):1162–1163.

Cooner, A., Shao, Y., and Campbell, J. (2016). Detection of Urban Damage Using Remote Sensing and Machine Learning Algorithms: Revisiting the 2010 Haiti Earthquake. *Remote Sensing*, 8(10):868.

Crook, J., Morris, F., Fitzpatrick, R. G. J., Peatman, S. C., Schwendike, J., Stein, T. H., Birch, C. E., Hardy, S., and Yang, G. (2024). Impact of the Madden–Julian oscillation and equatorial waves on tracked mesoscale convective systems over southeast Asia. *Quarterly Journal of the Royal Meteorological Society*, 150(760):1724–1751.

Da Silva, N. A. and Matthews, A. J. (2021). Impact of the Madden–Julian Oscillation on extreme precipitation over the western Maritime Continent and Southeast Asia. *Quarterly Journal of the Royal Meteorological Society*, 147(739):3434–3453.

Dai, A., Giorgi, F., and Trenberth, K. E. (1999). Observed and model-simulated diurnal cycles of precipitation over the contiguous United States. *Journal of Geophysical Research: Atmospheres*, 104(D6):6377–6402.

Das, S., Chakraborty, R., and Maitra, A. (2017). A random forest algorithm for nowcasting of intense precipitation events. *Advances in Space Research*, 60(6):1271–1282.

Davies, R. (2017). FloodList.

Dayem, K. E., Noone, D. C., and Molnar, P. (2007). Tropical western Pacific warm pool and maritime continent precipitation rates and their contrasting relationships with the Walker Circulation. *Journal of Geophysical Research*, 112(D6):D06101.

De Bruijn, J. A., De Moel, H., Jongman, B., De Ruiter, M. C., Wagemaker, J., and Aerts, J. C.

J. H. (2019). A global database of historic and real-time flood events based on social media. *Scientific Data*, 6(1):311.

De Ville, B. (2013). Decision trees. *WIREs Computational Statistics*, 5(6):448–455.

Dottori, F., Kalas, M., Salamon, P., Bianchi, A., Alfieri, L., and Feyen, L. (2017). An operational procedure for rapid flood risk assessment in Europe. *Natural Hazards and Earth System Sciences*, 17(7):1111–1126.

Environment Agency (2025). Environment Agency.

Ezugwu, A. E., Ho, Y.-S., Egwuche, O. S., Ekundayo, O. S., Van Der Merwe, A., Saha, A. K., and Pal, J. (2024). Classical Machine Learning: Seventy Years of Algorithmic Learning Evolution. *Data Intelligence*. Publisher: China Science Publishing & Media Ltd.

Falconer, R., Cobby, D., Smyth, P., Astle, G., Dent, J., and Golding, B. (2009). Pluvial flooding: new approaches in flood warning, mapping and risk management. *Journal of Flood Risk Management*, 2(3):198–208.

Feng, Z., Leung, L. R., Liu, N., Wang, J., Houze, R. A., Li, J., Hardin, J. C., Chen, D., and Guo, J. (2021). A Global High-Resolution Mesoscale Convective System Database Using Satellite-Derived Cloud Tops, Surface Precipitation, and Tracking. *Journal of Geophysical Research: Atmospheres*, 126(8):e2020JD034202.

Ferrett, S., Frame, T. H. A., Methven, J., Holloway, C. E., Webster, S., Stein, T. H., and Cafaro, C. (2021). Evaluating convection-permitting ensemble forecasts of precipitation over Southeast Asia. *Weather and Forecasting*.

Flack, D. L. A., Lehnert, M., Lean, H. W., and Willington, S. (2023). Characteristics of Diagnostics for Identifying Elevated Convection over the British Isles in a Convection-Allowing Model. *Weather and Forecasting*, 38(7):1079–1094.

Flack, D. L. A., Skinner, C. J., Hawkness-Smith, L., O'Donnell, G., Thompson, R. J., Waller, J. A., Chen, A. S., Moloney, J., Largeron, C., Xia, X., Blenkinsop, S., Champion, A. J., Perks, M. T., Quinn, N., and Speight, L. J. (2019). Recommendations for Improving Integration in National End-to-End Flood Forecasting Systems: An Overview of the FFIR (Flooding From Intense Rainfall) Programme. *Water*, 11(4):725.

Fleming, G., Rogers, D., Davies, P., Jacks, E., Milton, J. A., Honore, C., Lee, L. S., Bally, J.,

Zhihua, W., Tutis, V., and Goolaup, P. (2015). WMO Guidelines on Multi-hazard Impact-based Forecast and Warning Services. Technical report, World Meteorological Organization.

Fletcher, J. K., Diop, C. A., Adefisan, E., Ahiataku, M. A., Ansah, S. O., Birch, C. E., Burns, H. L., Clarke, S. J., Gacheru, J., James, T. D., Tuikong, C. K. N., Koros, D., Indasi, V. S., Lamptey, B. L., Lawal, K. A., Parker, D. J., Roberts, A. J., Stein, T. H. M., Visman, E., Warner, J., Woodhams, B. J., Youds, L. H., Ajayi, V. O., Bosire, E. N., Cafaro, C., Camara, C. A. T., Chanzu, B., Dione, C., Gitau, W., Groves, D., Groves, J., Hill, P. G., Ishiyaku, I., Klein, C. M., Marsham, J. H., Mutai, B. K., Ndiaye, P. N., Osei, M., Popoola, T. I., Talib, J., Taylor, C. M., and Walker, D. (2023). Tropical Africa's First Testbed for High-Impact Weather Forecasting and Nowcasting. *Bulletin of the American Meteorological Society*, 104(8):E1409–E1425.

Flood Forecasting Centre (2022). Flood Forecasting Centre.

Flood Forecasting Centre (2024). Flood Forecasting Centre.

FloodTags (2025). Flood Tags.

Foresti, L., Sideris, I. V., Nerini, D., Beusch, L., and Germann, U. (2019). Using a 10-Year Radar Archive for Nowcasting Precipitation Growth and Decay: A Probabilistic Machine Learning Approach. *Weather and Forecasting*, 34(5):1547–1569.

Fritz, S. and Wexler, H. (1960). CLOUD PICTURES FROM SATELLITE TIROS I*. *Monthly Weather Review*, 88(3):79–87.

Geen, R., Bordoni, S., Battisti, D. S., and Hui, K. (2020). Monsoons, ITCZs, and the Concept of the Global Monsoon. *Reviews of Geophysics*, 58(4):e2020RG000700.

Germann, U. and Zawadzki, I. (2002). Scale-Dependence of the Predictability of Precipitation from Continental Radar Images. Part I: Description of the Methodology. *Monthly Weather Review*, 130(12):2859–2873.

Giannini, A., Robertson, A. W., and Qian, J.-H. (2007). A role for tropical tropospheric temperature adjustment to El Niño–Southern Oscillation in the seasonality of monsoonal Indonesia precipitation predictability. *Journal of Geophysical Research*, 112(D16):D16110.

Gijben, M. and de Coning, E. (2017). Using Satellite and Lightning Data to Track Rapidly Developing Thunderstorms in Data Sparse Regions. *Atmosphere*, 8(12):67.

Golding, B. W. (1998). Nimrod: A system for generating automated very short range forecasts. *Meteorological Applications*, 5(1):1–16.

Goyal, S., Kumar, A., Mohapatra, M., Rathore, L. S., Dube, S. K., Saxena, R., and Giri, R. K. (2017). Satellite-based technique for nowcasting of thunderstorms over Indian region. *Journal of Earth System Science*, 126(6):79.

Green, D., Yu, D., Pattison, I., Wilby, R., Bosher, L., Patel, R., Thompson, P., Trowell, K., Draycon, J., Halse, M., Yang, L., and Ryley, T. (2017). City-scale accessibility of emergency responders operating during flood events. *Natural Hazards and Earth System Sciences*, 17(1):1–16.

Guerreiro, S., Glenis, V., Dawson, R., and Kilsby, C. (2017). Pluvial Flooding in European Cities—A Continental Approach to Urban Flood Modelling. *Water*, 9(4):296.

Hagos, S. (2010). Building Blocks of Tropical Diabatic Heating. *Journal of the Atmospheric Sciences*, 67(7):2341–2354. Publisher: American Meteorological Society.

Haiden, T., Kann, A., Wittmann, C., Pistotnik, G., Bica, B., and Gruber, C. (2011). The Integrated Nowcasting through Comprehensive Analysis (INCA) System and Its Validation over the Eastern Alpine Region. *Weather and Forecasting*, 26(2):166–183.

Han, D., Lee, J., Im, J., Sim, S., Lee, S., and Han, H. (2019). A Novel Framework of Detecting Convective Initiation Combining Automated Sampling, Machine Learning, and Repeated Model Tuning from Geostationary Satellite Data. *Remote Sensing*, 11(12):1454.

Han, L., Sun, J., Zhang, W., Xiu, Y., Feng, H., and Lin, Y. (2017). A machine learning nowcasting method based on real-time reanalysis data. *Journal of Geophysical Research: Atmospheres*, 122(7):4038–4051.

Han, L., Zhang, J., Chen, H., Zhang, W., and Yao, S. (2022). Toward the Predictability of a Radar-Based Nowcasting System for Different Precipitation Systems. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5.

Hand, W. H., Fox, N. I., and Collier, C. G. (2004). A study of twentieth-century extreme rainfall events in the United Kingdom with implications for forecasting. *Meteorological Applications*, 11(1):15–31.

Harjupa, W., Abdillah, M., Azura, A., Putranto, M., Marzuki, M., Nauval, F., Risyanto,

Saufina, E., Jumianti, N., and Fathrio, I. (2022). On the utilization of RDCA method for detecting and predicting the occurrence of heavy rainfall in Indonesia. *Remote Sensing Applications: Society and Environment*, 25:100681.

Harrison, D. L., Norman, K., Pierce, C., and Gaussiat, N. (2012). Radar products for hydrological applications in the UK. *Proceedings of the Institution of Civil Engineers - Water Management*, 165(2):89–103.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049.

Hill, P. G., Stein, T. H., Roberts, A. J., Fletcher, J. K., Marsham, J. H., and Groves, J. (2020). How skilful are Nowcasting Satellite Applications Facility products for tropical Africa? *Meteorological Applications*, 27(6).

Hollis, D., McCarthy, M., Kendon, M., Legg, T., and Simpson, I. (2019). HadUK-Grid—A new UK dataset of gridded climate observations. *Geoscience Data Journal*, 6(2):151–159.

Holmlund, K., Counet, P., Fadrique, F. M., Schmid, A., Bojkov, B., Munro, R., Grandell, J., and Obligis, E. (2025). The EUMETSAT satellite programmes and data services. *Journal of the European Meteorological Society*, 2:100005.

Horn, B. K. and Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T. (2014). The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society*, 95(5):701–722.

Houze, R. A. (2004). Mesoscale convective systems. *Reviews of Geophysics*, 42(4):2004RG000150.

Hu, G., Dance, S. L., Fowler, A., Simonin, D., Waller, J., Auligne, T., Healy, S., Hotta, D., Löhnert, U., Miyoshi, T., Privé, N. C., Stiller, O., Wang, X., and Weissmann, M. (2025). On methods for assessment of the value of observations in convection-permitting data assimilation and numerical weather forecasting. *Quarterly Journal of the Royal Meteorological Society*, 151(768):e4933.

Hurford, A. P., Priest, S. J., Parker, D. J., and Lumbroso, D. M. (2012). The effectiveness of extreme rainfall alerts in predicting surface water flooding in England and Wales. *International Journal of Climatology*, 32(11):1768–1774.

Imhoff, R. O., Brauer, C. C., Overeem, A., Weerts, A. H., and Uijlenhoet, R. (2020). Spatial and Temporal Evaluation of Radar Rainfall Nowcasting Techniques on 1,533 Events. *Water Resources Research*, 56(8).

Islas-Flores, D. and Magaldi, A. (2024). Assessing Nowcast Models in the Central Mexico Region Using Radar and GOES-16 Satellite Data. *Atmosphere*, 15(2):152.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, Honolulu, HI. IEEE.

J. Slingo, P. Inness, R. Neale, S. Woolnough, and G. Yang (2003). Scale interactions on diurnal toseasonal timescales and their relevanceto model systematic errors. *Annals of Geophysics*, 46(1).

Javelle, P., Organde, D., Demargne, J., Saint-Martin, C., De Saint-Aubin, C., Garandeau, L., and Janet, B. (2016). Setting up a French national flash flood warning system for ungauged catchments based on the AIGA method. *E3S Web of Conferences*, 7:18010.

Jianbo Shi and Tomasi (1994). Good features to track. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*, pages 593–600, Seattle, WA, USA. IEEE Comput. Soc. Press.

Jones, M. R., Fowler, H. J., Kilsby, C. G., and Blenkinsop, S. (2013). An assessment of changes in seasonal and annual extreme rainfall in the UK between 1961 and 2009. *International Journal of Climatology*, 33(5):1178–1194.

Jury, M. R. (2024). Structure, modulation and impacts of southeastern Africa's diurnal convection. *Theoretical and Applied Climatology*, 155(6):4851–4862.

Kikuchi, K. and Wang, B. (2008). Diurnal Precipitation Regimes in the Global Tropics*. *Journal of Climate*, 21(11):2680–2696.

Klein, C., Belušić, D., and Taylor, C. M. (2018). Wavelet Scale Analysis of Mesoscale Convective Systems for Detecting Deep Convection From Infrared Imagery. *Journal of Geophysical Research: Atmospheres*, 123(6):3035–3050.

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., Hatfield, S., Battaglia, P., Sanchez-Gonzalez, A., Willson, M., Brenner, M. P., and Hoyer, S. (2024). Neural general circulation models for weather and climate. *Nature*, 632(8027):1060–1066.

Krogh, A. (2008). What are artificial neural networks? *Nature Biotechnology*, 26(2):195–197.

La, T.-V., Nguyen, T. H., Matgen, P., and Chini, M. (2024). Early Flood Warning Using Satellite-Derived Convective System and Precipitation Data - A Retrospective Case Study of Central Vietnam. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 3752–3756, Athens, Greece. IEEE.

Lagerquist, R., Stewart, J., Ebert-Uphoff, I., and Christina, K. (2021). Using Deep Learning to Nowcast the Spatial Coverage of Convection from Himawari-8 Satellite Data. *Monthly Weather Review*, 149(12):3897–3921.

Lakshmanan, V., Smith, T., Stumpf, G., and Hondl, K. (2007). The Warning Decision Support System–Integrated Information. *Weather and Forecasting*, 22(3):596–612.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P. (2023). GraphCast: Learning skillful medium-range global weather forecasting. arXiv:2212.12794 [cs].

Laroche, S. and Zawadzki, I. (1994). A Variational Analysis Method for Retrieval of Three-Dimensional Wind Field from Single-Doppler Radar Data. *Journal of the Atmospheric Sciences*, 51(18):2664–2682.

Lean, H. W., Clark, P. A., Dixon, M., Roberts, N. M., Fitch, A., Forbes, R., and Halliwell,

C. (2008). Characteristics of High-Resolution Versions of the Met Office Unified Model for Forecasting Convection over the United Kingdom. *Monthly Weather Review*, 136(9):3408–3424.

Lebedev, V., Ivashkin, V., Rudenko, I., Ganshin, A., Molchanov, A., Ovcharenko, S., Grokhovetskiy, R., Bushmarinov, I., and Solomentsev, D. (2019). Precipitation Nowcasting with Satellite Imagery. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2680–2688, Anchorage AK USA. ACM.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Lee, S., Han, H., Im, J., Jang, E., and Lee, M.-I. (2017). Detection of deterministic and probabilistic convection initiation using Himawari-8 Advanced Himawari Imager data. *Atmospheric Measurement Techniques*, 10(5):1859–1874.

Lewis, M. W. and Gray, S. L. (2010). Categorisation of synoptic environments associated with mesoscale convective systems over the UK. *Atmospheric Research*, 97(1-2):194–213.

Li, P. and Lai, E. S. (2004). Short-range quantitative precipitation forecasting in Hong Kong. *Journal of Hydrology*, 288(1-2):189–209.

Line, W. E., Schmit, T. J., Lindsey, D. T., and Goodman, S. J. (2016). Use of Geostationary Super Rapid Scan Satellite Imagery by the Storm Prediction Center*. *Weather and Forecasting*, 31(2):483–494.

Liu, X., Chen, H., Han, L., and Ge, Y. (2022). A Machine Learning Approach for Convective Initiation Detection Using Multi-source Data. In *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 6518–6521, Kuala Lumpur, Malaysia. IEEE.

Lo, J. C.-F. and Orton, T. (2016). The general features of tropical Sumatra Squalls. 71(7):4.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. arXiv:1411.4038 [cs].

Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20(2):130–141.

Lucas, B. D. and Kanade, T. (1981). An Iterative Image Registration Technique with an Application to Stereo Vision. page 10.

Machado, L. A. T. and Laurent, H. (2004). The Convective System Area Expansion over Amazonia and Its Relationships with Convective System Life Duration and High-Level Wind Divergence. *Monthly Weather Review*, 132(3):714–725.

Madden, R. A. and Julian, P. R. (1971). Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific. *Journal of the Atmospheric Sciences*, 28(5):702–708.

Marcos (2015). NWC SAF convective precipitation product from MSG: A new day-time method based on cloud top physical properties. *Tethys, Journal of Weather and Climate of the Western Mediterranean*.

Masunaga, H. (2009). A 9-season TRMM Observation of the Austral Summer MJO and Low-frequency Equatorial Waves. *Journal of the Meteorological Society of Japan. Ser. II*, 87A:295–315.

Matsuno, T. (1966). Quasi-Geostrophic Motions in the Equatorial Area. *Journal of the Meteorological Society of Japan. Ser. II*, 44(1):25–43.

Maybee, B., Birch, C. E., Böing, S. J., Willis, T., Speight, L., Porson, A. N., Pilling, C., Shelton, K. L., and Trigg, M. A. (2024). FOREWARNS: development and multifaceted verification of enhanced regional-scale surface water flood forecasts. *Natural Hazards and Earth System Sciences*, 24(4):1415–1436.

McGovern, A., Chase, R. J., Flora, M., Gagne, D. J., Lagerquist, R., Potvin, C. K., Snook, N., and Loken, E. (2023). A Review of Machine Learning for Convective Weather. *Artificial Intelligence for the Earth Systems*, 2(3):e220077.

Mecikalski, J. R., Sandmæl, T. N., Murillo, E. M., Homeyer, C. R., Bedka, K. M., Apke, J. M., and Jewett, C. P. (2021). Random Forest Model to Assess Predictor Importance and Nowcast Severe Storms using High-Resolution Radar–GOES Satellite–Lightning Observations. *Monthly Weather Review*.

Mecikalski, J. R., Williams, J. K., Jewett, C. P., Ahijevych, D., LeRoy, A., and Walker, J. R. (2015). Probabilistic 0–1-h Convective Initiation Nowcasts that Combine Geostationary Satel-

lite Observations and Numerical Weather Prediction Model Data. *Journal of Applied Meteorology and Climatology*, 54(5):1039–1059.

Merk, D. and Zinner, T. (2013). Detection of convective initiation using Meteosat SEVIRI: implementation in and verification with the tracking and nowcasting algorithm Cb-TRAM. *Atmospheric Measurement Techniques*, 6(8):1903–1918.

Milan, M., Macpherson, B., Tubbs, R., Dow, G., Inverarity, G., Mittermaier, M., Halloran, G., Kelly, G., Li, D., Maycock, A., Payne, T., Piccolo, C., Stewart, L., and Wlasak, M. (2020). Hourly 4D-Var in the Met Office UKV operational forecast model. *Quarterly Journal of the Royal Meteorological Society*, 146(728):1281–1301. Publisher: Wiley.

Mori, S., Jun-Ichi, H., Tauhid, Y. I., Yamanaka, M. D., Okamoto, N., Murata, F., Sakurai, N., Hashiguchi, H., and Sribimawati, T. (2004). Diurnal Land–Sea Rainfall Peak Migration over Sumatera Island, Indonesian Maritime Continent, Observed by TRMM Satellite and Intensive Rawinsonde Soundings. *Monthly Weather Review*, 132(8):2021–2039.

Moseley, S. and Sandford, C. (2019). Blended probabilistic nowcasting with the IMPROVER postprocessing system.

Mueller, C., Saxen, T., Roberts, R., Wilson, J., Betancourt, T., Dettling, S., Oien, N., and Yee, J. (2003). NCAR Auto-Nowcast System. *Weather and Forecasting*, 18(4):545–561.

Murphy, A. H. and Epstein, E. S. (1989). Skill Scores and Correlation Coefficients in Model Verification. *Monthly Weather Review*, 117(3):572–582.

N. Takayabu, Y. (1994). Large-Scale Cloud Disturbances Associated with Equatorial Waves: Part II: Westward-Propagating Inertio-Gravity Waves. *Journal of the Meteorological Society of Japan. Ser. II*, 72(3):451–465.

Naha Biswas, A., Lee, Y. H., Tao Yeo, W., Low, W. C., Heh, D. Y., and Manandhar, S. (2024). Statistical Analysis of Atmospheric Delay Gradient and Rainfall Prediction in a Tropical Region. *Remote Sensing*, 16(22):4165.

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce (2015). NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive. Artwork Size: 565.24 Tbytes Pages: 565.24 Tbytes.

Notti, D., Giordan, D., Caló, F., Pepe, A., Zucca, F., and Galve, J. P. (2018). Potential and Limitations of Open Satellite Data for Flood Mapping. *Remote Sensing*, 10(11):1673.

Nowakowska, K. and Smith, J. (2024). Nowcasting: advances and applications. *Weather*, 79(11):378–379.

Organization, W. M. (2023). Guidelines for Satellite-based Nowcasting in Africa. Technical report, World Meteorological Organization.

O'Shea, K. and Nash, R. (2015). An Introduction to Convolutional Neural Networks. arXiv:1511.08458 [cs].

Palmer, T. N. (2006). Predictability of weather and climate: from theory to practice. In Palmer, T. and Hagedorn, R., editors, *Predictability of Weather and Climate*, pages 1–29. Cambridge University Press, 1 edition.

Pasche, O. C., Wider, J., Zhang, Z., Zscheischler, J., and Engelke, S. (2025). Validating Deep Learning Weather Forecast Models on Recent High-Impact Extreme Events. *Artificial Intelligence for the Earth Systems*, 4(1):e240033.

Peatman, S. C., Birch, C. E., Schwendike, J., Marsham, J. H., Dearden, C., Webster, S., Neely, R. R., and Matthews, A. J. (2023). The Role of Density Currents and Gravity Waves in the Offshore Propagation of Convection over Sumatra. *Monthly Weather Review*, 151(7):1757–1777.

Peatman, S. C., Matthews, A. J., and Stevens, D. P. (2014). Propagation of the Madden-Julian Oscillation through the Maritime Continent and scale interaction with the diurnal cycle of precipitation: MJO Propagation and Scale Interaction with the Diurnal Cycle. *Quarterly Journal of the Royal Meteorological Society*, 140(680):814–825.

Peatman, S. C., Schwendike, J., Birch, C. E., Marsham, J. H., Matthews, A. J., and Yang, G.-Y. (2021). A local-to-large scale view of Maritime Continent rainfall: control by ENSO, MJO and equatorial waves. *Journal of Climate*, pages 1–52.

Permana, D. S., Hutapea, T. D., Praja, A. S., Paski, J. A. I., Makmur, E. E. S., Haryoko, U., Umam, I. H., Saepudin, M., and Adriyanto, R. (2019). The Indonesia In-House Radar Integration System (InaRAISE) of Indonesian Agency for Meteorology Climatology and Geo-

physics (BMKG): Development, Constraint, and Progress. *IOP Conference Series: Earth and Environmental Science*, 303(1):012051.

Petty, G. W. (2006). *A first course in atmospheric radiation*. Sundog Pub, Madison, Wis, 2nd ed edition.

Pilling, C. (2016). New developments at the Flood Forecasting Centre: operations and flood risk guidance. pages 237–248, Venice, Italy.

Porson, A. N., Carr, J. M., Hagelin, S., Darvell, R., North, R., Walters, D., Mylne, K. R., Mittermaier, M. P., Willington, S., and Macpherson, B. (2020). Recent upgrades to the Met Office convective-scale ensemble: An hourly time-lagged 5-day ensemble. *Quarterly Journal of the Royal Meteorological Society*, 146(732):3245–3265.

Porson, A. N., Hagelin, S., Boyd, D. F., Roberts, N. M., North, R., Webster, S., and Lo, J. C. (2019). Extreme rainfall sensitivity in convective-scale ensemble modelling over Singapore. *Quarterly Journal of the Royal Meteorological Society*, 145(724):3004–3022.

Portalés-Julià, E., Mateo-García, G., Purcell, C., and Gómez-Chova, L. (2023). Global flood extent segmentation in optical satellite images. *Scientific Reports*, 13(1):20316.

Prein, A. F., Gobiet, A., Suklitsch, M., Truhetz, H., Awan, N. K., Keuler, K., and Georgievski, G. (2013). Added value of convection permitting seasonal simulations. *Climate Dynamics*, 41(9-10):2655–2677.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., Lipzig, N. P. M., and Leung, R. (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, 53(2):323–361.

Price, D., Hudson, K., Boyce, G., Schellekens, J., Moore, R. J., Clark, P., Harrison, T., Connolly, E., and Pilling, C. (2012). Operational use of a grid-based model for flood forecasting. *Proceedings of the Institution of Civil Engineers - Water Management*, 165(2):65–77.

Prudden, R., Adams, S., Kangin, D., Robinson, N., Ravuri, S., Mohamed, S., and Arribas, A. (2020). A review of radar-based nowcasting of precipitation and applicable machine learning techniques. *arXiv:2005.04988 [physics, stat]*. arXiv: 2005.04988.

Pulkkinen, S., Nerini, D., Pérez Hortal, A. A., Velasco-Forero, C., Seed, A., Germann, U., and

Foresti, L. (2019). Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1.0). *Geoscientific Model Development*, 12(10):4185–4219.

Qian, J.-H. (2008). Why Precipitation Is Mostly Concentrated over Islands in the Maritime Continent. *Journal of the Atmospheric Sciences*, 65(4):1428–1441.

Ramage, C. S. (1968). ROLE OF A TROPICAL "MARITIME CONTINENT" IN THE ATMOSPHERIC CIRCULATION [1]. *Monthly Weather Review*, 96(6):365–370.

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S., Prudden, R., Mandhane, A., Clark, A., Brock, A., Simonyan, K., Hadsell, R., Robinson, N., Clancy, E., Arribas, A., and Mohamed, S. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878):672–677.

Reen, B. P., Cai, H., and Raby, J. W. (2020). Preliminary Investigation of Assimilating Global Synthetic Weather Radar.

Ribeiro, B. Z., Machado, L. A. T., Huamán Ch., J. H., Biscaro, T. S., Freitas, E. D., Mozer, K. W., and Goodman, S. J. (2019). An Evaluation of the GOES-16 Rapid Scan for Nowcasting in Southeastern Brazil: Analysis of a Severe Hailstorm Case. *Weather and Forecasting*, 34(6):1829–1848.

Roberts, A. J., Fletcher, J. K., Groves, J., Marsham, J. H., Parker, D. J., Blyth, A. M., Adefisan, E. A., Ajayi, V. O., Barrette, R., de Coning, E., Dione, C., Diop, A., Foamouhoue, A. K., Gijben, M., Hill, P. G., Lawal, K. A., Mutemi, J., Padi, M., Popoola, T. I., Rípodas, P., Stein, T. H., and Woodhams, B. J. (2022). Nowcasting for Africa : advances, potential and value. *Weather*, 77(7):250–256.

Roberts, N. M. and Lean, H. W. (2008). Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Monthly Weather Review*, 136(1):78–97.

Roca, R., Fiolleau, T., and Bouniol, D. (2017). A Simple Model of the Life Cycle of Mesoscale Convective Systems Cloud Shield in the Tropics. *Journal of Climate*, 30(11):4283–4298.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Publisher: arXiv Version Number: 1.

Rouault, E., Warmerdam, F., Schwehr, K., Kiselev, A., Butler, H., Łoskot, M., Szekeres, T., Tourigny, E., Landa, M., Miara, I., Elliston, B., Chaitanya, K., Plesea, L., Morissette, D., Jolma, A., Dawson, N., Baston, D., de Stigter, C., and Miura, H. (2023). GDAL.

Rouault, M., Dieppois, B., Tim, N., Hünicke, B., and Zorita, E. (2024). Southern Africa Climate Over the Recent Decades: Description, Variability and Trends. In Von Maltitz, G. P., Midgley, G. F., Veitch, J., Brümmer, C., Rötter, R. P., Viehberg, F. A., and Veste, M., editors, *Sustainability of Southern African Ecosystems under Global Change*, volume 248, pages 149–168. Springer International Publishing, Cham. Series Title: Ecological Studies.

Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., and Aberman, K. (2023). DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. arXiv:2208.12242 [cs].

Ruthotto, L. and Haber, E. (2021). An Introduction to Deep Generative Modeling. arXiv:2103.05180 [cs].

Rüdisühli, S., Sprenger, M., Leutwyler, D., Schär, C., and Wernli, H. (2020). Attribution of precipitation to cyclones and fronts over Europe in a kilometer-scale regional climate simulation. *Weather and Climate Dynamics*, 1(2):675–699.

Saint-Martin, C., Fouchier, C., Javelle, P., Douvinet, J., and Vinet, F. (2016). Assessing the exposure to floods to estimate the risk of flood-related damage in French Mediterranean basins. *E3S Web of Conferences*, 7:04013.

Saltikoff, E., Friedrich, K., Soderholm, J., Lengfeld, K., Nelson, B., Becker, A., Hollmann, R., Urban, B., Heistermann, M., and Tassone, C. (2019). An Overview of Using Weather Radar for Climatological Studies: Successes, Challenges, and Potential. *Bulletin of the American Meteorological Society*, 100(9):1739–1752.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., and Ratier, A. (2002). Supplement to An Introduction to Meteosat Second Generation (MSG): SEVIRI CALIBRATION. *Bulletin of the American Meteorological Society*, 83(7):992–992.

Schmid, F., Wang, Y., and Harou, A. (2019). Nowcasting Guidelines – A Summary. *World Meteorological Organization*.

Schmit, T. J., Li, J., Li, J., Feltz, W. F., Gurka, J. J., Goldberg, M. D., and Schrab, K. J.

(2008). The GOES-R Advanced Baseline Imager and the Continuation of Current Sounder Products. *Journal of Applied Meteorology and Climatology*, 47(10):2696–2711.

Schumacher, C. and Houze, R. A. (2003). Stratiform Rain in the Tropics as Seen by the TRMM Precipitation Radar*. *Journal of Climate*, 16(11):1739–1756.

Schumacher, R. S. and Rasmussen, K. L. (2020). The formation, character and changing nature of mesoscale convective systems. *Nature Reviews Earth & Environment*, 1(6):300–314.

Seal, D., Arcucci, R., Rühling-Cachay, S., and Quilodrán-Casas, C. (2024). DYffCast: Regional Precipitation Nowcasting Using IMERG Satellite Data. A case study over South America. arXiv:2412.02723 [cs].

Seed, A. W. (2003). A Dynamic and Spatial Scaling Approach to Advection Forecasting. *Journal of Applied Meteorology*, 42(3):381–388.

Seed, A. W., Pierce, C. E., and Norman, K. (2013). Formulation and evaluation of a scale decomposition-based stochastic precipitation nowcast scheme: Formulation of a Scale Decomposition Nowcast Scheme. *Water Resources Research*, 49(10):6624–6641.

Sen Roy, S., Saha, S. B., Roy Bhowmik, S. K., and Kundu, P. K. (2014). Optimization of Nowcast Software WDSS-II for operational application over the Indian region. *Meteorology and Atmospheric Physics*, 124(3-4):143–166.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., and Woo, W.-c. (2015). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *arXiv:1506.04214 [cs]*. arXiv: 1506.04214.

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-k., and Woo, W.-c. (2017). Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model. Version Number: 2.

Sobajima, A. (2012). Rapidly Developing Cumulus Areas Derivation Algorithm Theoretical Basis Document.

Speight, L., Birch, C., Self, K., and Brown, S. (2025). Expert perspectives on the next generation of UK surface water flood warning services. *Weather*. Publisher: Wiley.

Speight, L., Cole, S., Moore, R., Pierce, C., Wright, B., Golding, B., Cranston, M., Tavendale,

A., Dhondia, J., and Ghimire, S. (2018). Developing surface water flood forecasting capabilities in Scotland: an operational pilot for the 2014 Commonwealth Games in Glasgow. *Journal of Flood Risk Management*, 11(S2).

Speight, L. J., Cranston, M. D., White, C. J., and Kelly, L. (2021). Operational and emerging capabilities for surface water flood forecasting. *WIREs Water*, 8(3):e1517.

Srivastava, K., Lau, S., Yeung, H., Cheng, T., Bhardwaj, R., Kannan, A., Bhowmik, S., and Singh, H. (2021). Use of SWIRLS nowcasting system for quantitative precipitation forecast using Indian DWR data. *MAUSAM*, 63(1):1–16.

Straub, K. H. and Kiladis, G. N. (2002). Observations of a Convectively Coupled Kelvin Wave in the Eastern Pacific ITCZ. *Journal of the Atmospheric Sciences*, 59(1):30–53.

Sun, F., Li, B., Min, M., and Qin, D. (2023). Toward a Deep-Learning-Network-Based Convective Weather Initiation Algorithm From the Joint Observations of Fengyun-4A Geostationary Satellite and Radar for 0–1h Nowcasting. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:3455–3468.

SWIFT, G. A. (2025). GCRF African SWIFT.

Taromideh, F., Santonastaso, G. F., and Greco, R. (2025). Rainfall nowcasting by integrating radar and rain gauge data with machine learning for Ischia Island, Italy. *Journal of Hydrology: Regional Studies*, 58:102273.

Taylor, C. M., De Jeu, R. A. M., Guichard, F., Harris, P. P., and Dorigo, W. A. (2012). Afternoon rain more likely over drier soils. *Nature*, 489(7416):423–426.

Taylor, C. M., Klein, C., Dione, C., Parker, D. J., Marsham, J., Abdoulahat Diop, C., Fletcher, J., Saidou Chaibou, A. A., Nafissa, D. B., Semeena, V. S., Cole, S. J., and Anderson, S. R. (2022). Nowcasting tracks of severe convective storms in West Africa from observations of land surface state. *Environmental Research Letters*, 17(3):034016.

Thu, N. V., Tri, D. Q., Hoa, B. T. K., Nguyen-Thi, H. A., Hung, N. V., Tuyet, Q. T. T., Van Nhat, N., and Pham, H. T. T. (2025). Application of optical flow technique to short-term rainfall forecast for some synoptic patterns in Vietnam. *Theoretical and Applied Climatology*, 156(1):1.

UK Parliament (2010). Flood Water Management Act 2010. Technical report.

Vesuviano, G. (2022). The FEH22 rainfall depth- duration-frequency (DDF) model. Technical report.

Vila, D. A., Machado, L. A. T., Laurent, H., and Velasco, I. (2008). Forecast and Tracking the Evolution of Cloud Clusters (ForTraCC) Using Satellite Infrared Imagery: Methodology and Validation. *Weather and Forecasting*, 23(2):233–245.

Vogel, P., Knippertz, P., Fink, A. H., Schlueter, A., and Gneiting, T. (2020). Skill of Global Raw and Postprocessed Ensemble Predictions of Rainfall in the Tropics. *Weather and Forecasting*, 35(6):2367–2385.

Wang, B. and Ding, Q. (2008). Global monsoon: Dominant mode of annual variation in the tropics. *Dynamics of Atmospheres and Oceans*, 44(3-4):165–183.

Wang, R., Fung, J. C. H., and Lau, A. K. H. (2024). Skillful Precipitation Nowcasting Using Physical-Driven Diffusion Networks. *Geophysical Research Letters*, 51(24):e2024GL110832.

Wheeler, M., Kiladis, G. N., and Webster, P. J. (2000). Large-Scale Dynamical Fields Associated with Convectively Coupled Equatorial Waves. *Journal of the Atmospheric Sciences*, 57(5):613–640.

Wheeler, M. C. and Hendon, H. H. (2004). An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction. *Monthly Weather Review*, 132(8):1917–1932.

Willems, P. (2016). Prototype system for nowcasting of urban flood risks.

Woodhams, B. J., Birch, C. E., Marsham, J. H., Bain, C. L., Roberts, N. M., and Boyd, D. F. A. (2018). What Is the Added Value of a Convection-Permitting Model for Forecasting Extreme Rainfall over Tropical East Africa? *Monthly Weather Review*, 146(9):2757–2780.

Yang, G.-Y., Hoskins, B., and Slingo, J. (2003). Convectively Coupled Equatorial Waves: A New Methodology for Identifying Wave Structures in Observational Data. *Journal of the Atmospheric Sciences*, 60(14):1637–1654.

Yang, G.-Y. and Slingo, J. (2001). The Diurnal Cycle in the Tropics. *Monthly Weather Review*, 129(4):784–801.

Yang, J., Zhang, Z., Wei, C., Lu, F., and Guo, Q. (2017). Introducing the New Generation of

Chinese Geostationary Weather Satellites, Fengyun-4. *Bulletin of the American Meteorological Society*, 98(8):1637–1658.

Yu, D. and Coulthard, T. J. (2015). Evaluating the importance of catchment hydrological parameters for urban surface water flood modelling using a simple hydro-inundation model. *Journal of Hydrology*, 524:385–400.

Zhan, W. and Datta, A. (2024). Neural Networks for Geospatial Data. *Journal of the American Statistical Association*, pages 1–21.

Zhang, C. (2005). Madden-Julian Oscillation: MADDEN-JULIAN OSCILLATION. *Reviews of Geophysics*, 43(2).

Zhang, S. and Wang, B. (2008). Global summer monsoon rainy seasons. *International Journal of Climatology*, 28(12):1563–1578.

Zhang, Y., Long, M., Chen, K., Xing, L., Jin, R., Jordan, M. I., and Wang, J. (2023a). Skilful nowcasting of extreme precipitation with NowcastNet. *Nature*, 619(7970):526–532.

Zhang, Y., Zheng, X., Li, X., Lyu, J., and Zhao, L. (2023b). Evaluation of the GPM-IMERG V06 Final Run Products for Monthly/Annual Precipitation under the Complex Climatic and Topographic Conditions of China. *Journal of Applied Meteorology and Climatology*, 62(8):929–946.

Zhao, Z., Dong, X., Wang, Y., and Hu, C. (2024). Advancing Realistic Precipitation Nowcasting With a Spatiotemporal Transformer-Based Denoising Diffusion Model. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15.

Zhu, H., Wheeler, M. C., Sobel, A. H., and Hudson, D. (2014). Seamless Precipitation Prediction Skill in the Tropics and Extratropics from a Global Model. *Monthly Weather Review*, 142(4):1556–1569.

Zhu, Z., Gu, J., Xu, B., and Shi, C. (2024). Characterization of Himawari-8/AHI to Himawari-9/AHI infrared observations continuity. *International Journal of Remote Sensing*, 45(1):121–142.

Zinner, T., Mannstein, H., and Tafferner, A. (2008). Cb-TRAM: Tracking and monitoring severe convection from onset over rapid development to mature phase using multi-channel Meteosat-8 SEVIRI data. *Meteorology and Atmospheric Physics*, 101(3-4):191–210.

Figure A.2: 25th September 2025