Received: 26 March 2021 | Revised: 5 October 2021 | Accepted: 12 November 2021

DOI: 10.1002/met.2038

Meteorological Applications
Science and Technology for Weather and Climate
Open Access
RMetS

RESEARCH ARTICLE

Convective-scale numerical weather prediction and warnings over Lake Victoria: Part I—Evaluating a lightning diagnostic

Marion Mittermaier[1] | **Jonathan Wilkinson**[1] | **Gabriella Csima**[1] | **Steven Goodman**[2] | **Katrina Virts**[3]

[1]Met Office, Exeter, UK

[2]Thunderbolt Analytics, Huntsville, Alabama, USA

[3]The University of Alabama in Huntsville, Huntsville, Alabama, USA

Correspondence
Marion Mittermaier, Met Office, FitzRoy Road, Exeter, EX1 3PB, UK.
Email: marion.mittermaier@metoffice.gov.uk

Funding information
Steven Goodman was supported in part by NASA Grant 80NSSC18K1655 and Met Office Subcontract # P105101. Katrina Virts was supported by the NASA Post-Doctoral Program and the GOES-R Satellite Program. This research was conducted with financial support from UKAID from the UK government Foreign and Commonwealth Development Office (FCDO) under the WISER (Weather and Climate Information Services for Africa) Programme through the HIGHWAY project.

Abstract

Recently a new lightning diagnostic was added to convective-scale configurations of the Met Office Unified Model (UM). The characteristics and skill of the hourly total 'flash origin density' in the 4.4 km Tropical Africa UM were evaluated against Earth Network (EN) lightning observations specifically deployed in the greater Lake Victoria region as part of the HIGHWeather impact lAke sYstem (HIGHWAY) project. Lightning poses a significant risk to fishermen on Lake Victoria, and better forecasts over the lake would be advantageous for the development of better warning systems. The model lightning density was evaluated in two ways: a classical gridded categorical data analysis and the spatial *coverage–distance–intensity* method. Given the spatial sparseness of the observation type, Gaussian kernel dressing was applied for the latter, to increase the horizontal 'footprint', partly to reflect that lightning can travel horizontally for more than 10 km, and to reduce the representativeness mismatch between the lightning observations and the model forecast lightning density. Considering hourly forecasts, lightning shows a marked diurnal, geographical and seasonal variation over to the west and east of the lake. The model is not producing enough lightning flashes over the lake, nor does it produce enough spatially, especially overnight. The forecast location of lightning shows little skill, with timing differences in the peak, especially over the lake. For warnings, the removal of timing offsets through the creation of a 'maximum-in-the-day' field is considerably less biased and more skilful than using the hourly flash densities.

KEYWORDS

forecasting, hazards, lightning, remote sensing, severe weather, short-range

This article is published with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2022 Crown copyright. *Meteorological Applications* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1 | INTRODUCTION

Forecasting thunderstorms in the deep tropics is challenging for numerical weather prediction (NWP) models irrespective of resolution (e.g., Huang et al., 2019; Sun et al., 2020), yet this is what we want to do because thunderstorms are a multi-hazard phenomenon, which poses multiple risks to people and their livelihoods: lightning, torrential rain leading to flash flooding, strong winds and gusts are all common hazards associated with these storms. Recently the HIGHWeather impact lAke sYstem (HIGHWAY) project in the Lake Victoria region (Figure 1) has been a significant focus for understanding how warnings of thunderstorm hazards can be improved to reduce the risk to the people having to live with thunderstorms. NWP output, especially from convective-scale models, is seen to be an important part of this, and this paper aims to specifically explore the characteristics of NWP lightning forecasts from a modeller's perspective.

Lake Victoria is the largest freshwater body in Africa, with a surface area of ~50,000–60,000 km^2, located 1134 m ASL. It acts as a focal point for thunderstorm development. Lake storms are often associated with strong winds, which in turn lead to large waves, both significant hazards for the local population. The region is characterized by some of the strongest storms globally (e.g., Zipser et al., 2006), which is due to the region's geography. Immediately to the west and north-west of the lake, the terrain is relatively flat, while to the east it is very mountainous. The lake and the local wind circulations it produces provide an important driver for the local climate, which is markedly different compared with other equatorial locations. For example, it is much drier, with a mean maximum rainfall of 500–1500 mm per year. This is ascribed to the divergent flow resulting from the north-east and south-west monsoons (e.g., Trewartha, 1961) as well as the inter-tropical convergence zone (ITCZ) being interrupted and modified due to the high orography of the region (Griffiths, 1972; Nicholson, 1996).

There are two rainy seasons, the 'long' (April–May) and 'short' or 'light' rains (November–December), the exception being the lake itself, which has rainfall all year round. The 'short rains' can be strongly affected by El Niño/Southern Oscillation (ENSO) (e.g., Ogallo, 1988), while the 'long rains' are characterized by strong thunderstorms and torrential rainfall.

Virts and Goodman (2020) have found that the vast majority (about 85%) of lightning storm clusters producing more than 1000 strokes over Lake Victoria initiate in situ over the lake, but 30%–40% of the more extreme clusters initiated over land, mostly to the east of the lake, before propagating *westwards* across the lake in the prevailing easterly flow (e.g., Anyah et al., 2006; Ba & Nicholson, 1998; Song et al., 2004). The highest total lightning density in this region (about 450 strokes km^{-2} year^{-1}) is over Lake Victoria, where thunderstorms are an almost daily occurrence. Seasonal shifts in the ITCZ produce semi-annual lightning maxima over the lake. Diurnally, solar heating and lake and valley breezes produce daytime lightning maxima north and east of the lake, while at night the peak lightning density propagates southwestward across the lake. These findings expand on the pioneering results by Griffiths (1972), who noted that over the lake the high maximum of convection (from a precipitation perspective) happened at different times of day depending on the season, with months reporting more than 200 mm rainfall showing very little diurnal variation. Griffiths (1972) also noted the observed minimum or lull in convection around midday local time (09 UTC). More recently, Yin et al. (2000) showed the presence of an east–west split in the diurnal cycle over the lake: an early morning peak in convection in the west and north-west, and a late afternoon peak over the eastern lake.

Focusing on the diurnal cycle of lightning instead of precipitation, Virts and Goodman (2020) found that the most prolific lightning-producing thunderstorms initiate between 11:00 and 14:00 local time (LT) (08–11 UTC). Initiation times of prolific Lake Victoria storms exhibit a bimodal seasonal cycle: equinox-season thunderstorms initiate most frequently between 22:00 and 04:00 LT (19–01 UTC), while solstice-season thunderstorms initiate most frequently from 05:00 to 08:00 LT (02–05 UTC),

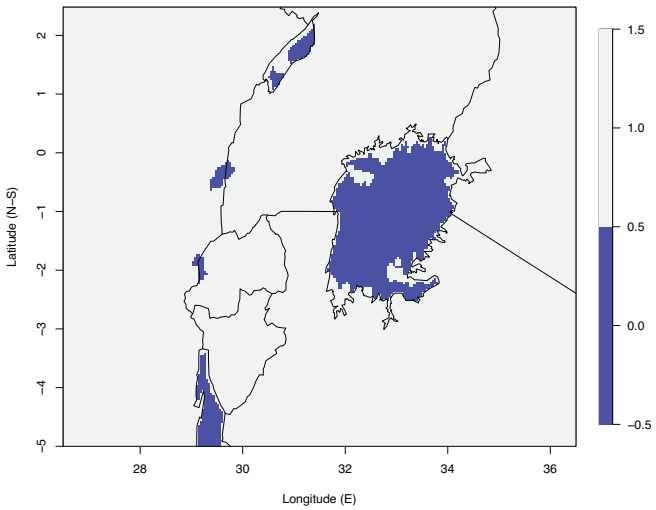

FIGURE 1 Map of the study region showing the land–sea mask defined within the model ancillary fields. The model resolution is 0.04° (4.4 km) with the domain stretching from 26.5° E to 36.5° E and 5° S to 2.5° N providing a 250 × 187 grid squares. Note a map of the terrain is available in part 2 of this paper (Mittermaier et al., 2021, fig. 2)

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

more than 12 h after the afternoon convective peak over land. Solar heating and the associated lake and valley breezes lead to an afternoon maximum in deep moist convection over land, with lightning peaking around 13:00–16:00 LT (10–13 UTC) followed by peak precipitation around 15:00–18:00 LT (12–15 UTC). At night, a deep convective maximum shift from northeast to southwest over the lake is observed, frequently including upscale growth into mesoscale convective systems.

Overall, the pattern of convection, precipitation and lightning in the region is highly repetitive, and while there are subtle differences in intensity and location as a function of time of year, using observed persistence (in the absence of NWP) as a starting point for thunderstorm forecasting in the region would seem a very sensible strategy. Comparing the performance of NWP forecasts to persistence could also make the NWP forecasts look like they add little value, as persistence under certain conditions could be rather skillful and tough to beat (e.g., Mittermaier, 2008).

Indeed, Lafore et al. (2017) note that forecasters rely on their local knowledge and experience of interpreting larger-scale synoptic drivers (available from global models) to infer whether conditions for thunderstorm formation are favourable or likely. Several recent studies have considered the value of convective-scale models (over global models) for forecasting precipitation in this region. Chamberlain et al. (2014) show that a 4 km version of the Met Office Unified Model (UM) produced better guidance for issuing warnings than the equivalent global model at that time, producing fewer missed storms but introducing many false alarms. Woodhams et al. (2018) found that for precipitation on the *daily* timescale, both the global- and convective-scale models can outperform a 24 h persistence forecast, with the convective-scale model exhibiting greater skill than the global model, with a better diurnal cycle of convection. Finally, Hanley et al. (2020) provide a model-oriented evaluation of the UM configurations relevant to the HIGHWAY project, showing that the UM tends to produce too much precipitation and too many small storms. They also found the modelled onset time of daytime precipitation over land was a good fit to the observed, but night-time precipitation was triggered late over Lake Victoria with precipitation amounts under-estimated. Hanley et al. (2020) suggest this is because some nocturnal lake storms are simply not initiated by the model.

The availability of lightning flash density forecasts from convective-scale NWP is still relatively new. Most of the evaluation thus far has been related to parameterization development (e.g., McCaul et al., 2009). Early verification studies (e.g., Federico et al., 2014; Giannaros et al., 2017; Groenemeijer et al., 2007; Lynn et al., 2012) included some categorical statistics, while Lynn et al. (2015) and Lynn (2017) also experimented with neighbourhood verification methods (see, e.g., Ebert, 2008 for an introduction to this topic). Further objective verification and diagnostic studies, primarily related to comparing different electrification parameterization schemes, have been documented more recently (e.g., Dafis et al., 2018; Mohan et al., 2021; Qian & Wang, 2021). Wilkinson (2017) introduced a new technique specifically for evaluating deterministic lightning forecasts. This method (which has been adapted slightly) is applied here, alongside a selection of traditional categorical metrics. Furthermore, a novel way of pre-processing the gridded lightning observations to increase their footprint is also demonstrated in this paper, which documents the evaluation of the 'lightning origin flash density' forecasts from the 4.4 km Tropical African configuration of the Met Office UM. Forecasts are evaluated for the period April–October 2019 (192 days) to understand the diurnal cycle, biases and skill. The paper is organized as follows: the data sets are described in Section 2; the methodology is described in Section 3. The results follow in Section 4. Conclusions and recommendations are provided in Section 5.

2 | DATA SETS

2.1 | Model description

The Tropical Africa model is a 4.4 km convective-scale configuration of the Met Office UM. The UM is non-hydrostatic utilizing a semi-implicit, semi-Lagrangian numerical scheme (Davies et al., 2005; Wood et al., 2014). The model uses the following parameterization schemes: surface (Best et al., 2011); mixed-phase cloud microphysics (Wilson & Ballard, 1999); and planetary boundary layer (PBL: Lock et al., 2000). The convection scheme (Gregory & Rowntree, 1990) is switched off for grid lengths of 4.4 km and below. High-resolution versions of the UM use a stability-dependent Smagorinsky-type subgrid turbulence scheme, which is blended with the non-local PBL scheme (Boutle et al., 2014). This allows the model to transition from unresolved to resolved turbulence in a gradual way.

This configuration has 80 vertical levels and uses the Regional Atmosphere and Land version 1.0 Tropical (RAL1-T) configuration (Bush et al., 2019). In this configuration convection is treated explicitly, which means that the CAPE-dependent closure scheme is switched off. RAL1-T also uses a prognostic large-scale cloud scheme (PC2, Wilson et al., 2008). Only a fraction of the full Tropical Africa model domain is of interest in this paper and is shown in Figure 1 as the land–sea mask, providing a 250 × 187 grid point domain, which straddles the Equator.

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The regional model is one-way nested inside the Met Office Global Model, which was based on the Global Atmosphere version 6.1 (GA6.1) configuration (Walters et al., 2011) at the time. A hybrid incremental 4D-Var data assimilation system (Rawlins et al., 2007) produces a new global analysis every 6 h. There is no direct data assimilation for this regional configuration. The Tropical Africa model has been operational since March 2019 and is initialized twice a day from the GM analysis at 06 and 18 UTC producing a 48 h forecast with 3-hourly updated lateral boundary conditions (LBCs) from the GM. Sea- and lake-surface temperatures are updated daily using the operational sea surface temperature and sea ice analysis (OSTIA: Fiedler et al., 2014).

2.2 | Lightning diagnostic definition

Following similar limited area configurations of the Met Office UM (Bush et al., 2019; Stratton et al., 2018), the microphysics scheme is based on Wilson and Ballard (1999), but with the addition of prognostic variables for rain and graupel. A single ice category is used to represent both cloud ice and snow, while the graupel category represents all ice that is denser than snow, including hail. Charge separation in thunderclouds is generally believed to be as a result of non-inductive charging between ice and graupel (e.g., Emersic & Saunders, 2010; Saunders et al., 2007), and the addition of the prognostic graupel category allows the model to forecast lightning flash rate. The lightning scheme in the model is based on McCaul et al. (2009), who compared observed lightning flash rate from a lightning mapping array against properties from a 2 km grid convective-scale model over Alabama. They link the total lightning flash rate as a weighted function of two quantities ($F1$ and $F2$) as follows:

$$F = r_1 F1 + r_2 F2, \tag{1}$$

where r_1 and r_2 are weighting factors: $r_1 = 0.95$ flash rate per minute times s m^{-1} and $r_2 = 0.05$ flash rate per minute times (m*m) kg^{-1}. In the UM, $F1$ is expressed as the flash rate per minute as a result of the upward flux of graupel at the −15°C level:

$$F1 = 0.042\left[wq_g\right]_{-15^\circ\mathrm{C}}, \tag{2}$$

where w is the model vertical velocity at the minus −15 °C level (m s^{-1}) and q_g is the mass mixing ratio of graupel at the same isotherm (as kg of graupel per kg of air). The second term is due to the total ice water path, or vertically integrated ice in the model column:

$$F2 = 0.02 \int_0^{TOA} \rho\left(q_i + q_g\right) dz. \tag{3}$$

Here, $F2$ is again taken in flashes per minute, while ρ is the density of air (kg m^{-3}), dz is the thickness of the model layer (m) and q_i is the mixing ratio of the combined ice and snow category (kg of ice per kg of dry air). Lightning flash rates are only calculated where the vertical integral of graupel in the model column is greater than 0.2 kg m^{-3}. On each timestep, the flash rate is multiplied by the length of the model timestep (100 s) and the total number of lightning flashes in that model grid column is computed. The model then discharges the integer component of the lightning flashes on that timestep. Any remaining flash is then passed through the following timestep and will contribute to the lightning flashes on the following timestep, provided the vertical integral of graupel is above the 0.2 kg m^{-3} threshold. Should this not be the case, the unused flash is discarded. The resulting model output is a two-dimensional field of flash origin densities in each hour of the model run. No discrimination is made between intra-cloud and cloud-to-ground lightning in either the original McCaul et al. (2009) scheme or in its representation in the UM. The term 'flash origin densities' is used to indicate the source of the lightning flashes, and it is recognized that lightning flashes may propagate from one grid box to the next; however, this specific process is not represented in the model. A revised version of the lightning scheme was recently published, although this has not been implemented in the UM as yet (McCaul et al., 2020).

In this study, the forecasts initialized at 18 UTC were evaluated. The full forecast is 54 h long, but in line with part 2 of this study, the verification is focused on the definition of 'Day 1', which spans the 24 h period between 06 and 06 UTC, in line with the issuance of warnings (see companion paper Mittermaier et al., 2021), covered by forecast ranges of $t + 13$ h to $t + 36$ h. Forecasts were considered in two ways: hourly fields and as a 'maximum-in-the-day' field, extracting the maximum number of lightning flashes in each grid box within a 24-h period from the 24 underpinning hourly fields.

2.3 | Lightning observations

Beginning in 2014, Earth Networks Global Lightning Network (ENGLN) established a regional network of lightning sensors around Lake Victoria for observing total lightning (in-cloud (IC) and cloud-to-ground (CG)) using the time-of-arrival (TOA) technique (Nag et al., 2015).

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

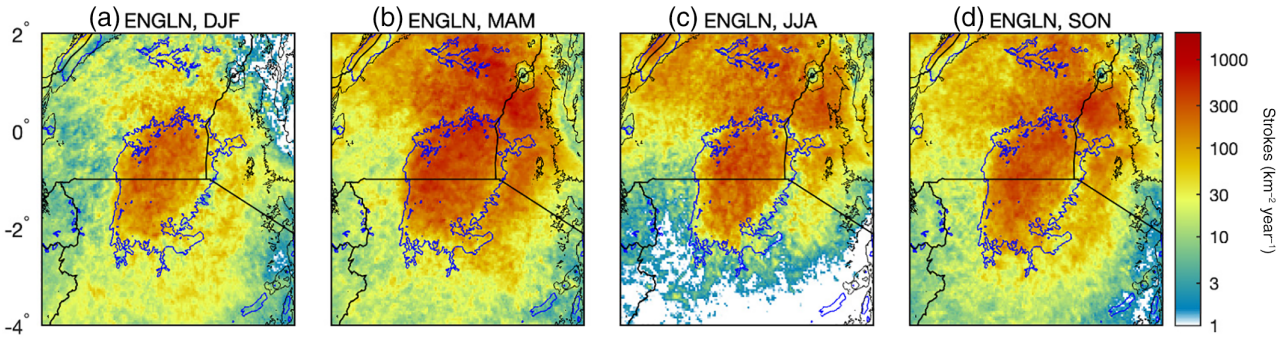

FIGURE 2 ENGLN average seasonal lightning strokes (km^{-2} year^{-1}) for the 3-year period March 2017–February 2020. DJF, December–February; MAM, March–May; JJA, June–August; SON, September–November

Prior climatological lightning studies in this region used either ground-based networks of widely spaced receivers for detecting cloud-to-ground lightning strokes (Holle & Murphy, 2017; Virts et al., 2013) or low Earth-orbiting optical imagers on satellites that detected total lightning only briefly during the satellite overpass (Albrecht et al., 2016).

The EN data were made available to researchers and forecasters for the duration of the HIGHWAY project (Virts & Goodman, 2020). The NASA/UAH Global Hydrology Resource Center Distributed Active Archive Center (GHRC DAAC) continually collects the near real-time streaming lightning data directly from Earth Networks encompassing the Lake Victoria Basin in East Africa (https://ghrc.nsstc.nasa.gov/home/field-campaigns/highway). The data for the HIGHWAY project are ingested every 5 min within a subset region described by a bounding box around Lake Victoria extending from 12° S to 4° N latitude and from 26° W to 40° E longitude. The proximity of receiving stations surrounding the lake adds to the improved detection of cloud flashes (and thus total lightning) to the system's ability to detect ground strokes. The data were uploaded directly to the Met Office in the United Kingdom.

Owing to the high density of ENGLN stations surrounding Lake Victoria, it was found that ENGLN is the best network for total lightning detection and mapping for storms initiating and propagating across the lake. Monthly ENGLN lightning climatologies and peak lightning cluster maps over Lake Victoria were produced throughout the HIGHWAY project. The seasonal lightning climatology is shown in Figure 2 during the full HIGHWAY project study period of March 2017–February 2020. Nearly 100 million lightning strokes were observed by ENGLN during the 3-year period, which is considered typical for this region as reported in the recent paper by Virts and Goodman (2020). The high ratio of intra-cloud (IC) to cloud-to-ground (CG) lightning observed by ENGLN in this region, calculated at 14.1 during the study period (Virts & Goodman, 2020), is characteristic of the more intense storms having stronger updrafts and deeper precipitation cores than ordinary air mass thunderstorms (Cecil et al., 2005). A typical airmass storm will have IC:CG or z-ratio (Prentice & Mackerras, 1977) of about 4:1. Having an awareness of the high proportion of CG lightning strokes is important for the interpretation of model output to produce severe weather warnings in the region, a topic that is explored in Mittermaier et al. (2021). Looking further to the future, Meteosat Third Generation (MTG) will carry a new Lightning Imager (LI) covering the whole of Africa. This study could be seen as preparatory for the MTG LI and its use in NWP.

Finally, it is important to emphasize that the model provides forecasts of total lightning flash density while the ENGLN observations are in units of strokes. A lightning *flash* consists of one or more *strokes*. The average number of strokes per flash can vary considerably (see, e.g., Rakov, 2016). For comparing the model lightning flash density, the ENGLN individual strokes were aggregated onto the model 4.4 km grid (domain shown in Figure 1) to create hourly gridded (frequency) fields. A mean of 2.1 strokes per flash was derived from the HIGHWAY ENGLN dataset and used to convert the number of strokes to flashes. A maximum-in-the-day field was also created to span the period from 06 to 06 UTC. Lightning strokes are extremely localized in nature and need to be post-processed to account for the fact that (a) lightning can travel more than 10 km in distance, which (b) introduce sufficient uncertainties to the location information, and which warrants some post-processing to account for the huge differences in representativity between the forecast lightning density diagnostic and the observed lightning strokes. Converting strokes to flashes and aggregating onto a grid are two of the mechanisms that can be used. Another new, additional methodology to reduce representativeness mismatches is described in Section 3.

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3 | VERIFICATION METHODOLOGY

3.1 | Spatial processing of gridded lightning observations for traditional categorical analysis

As explained in the previous section, even when accumulating lightning occurrence onto a grid, the grid can still be very sparsely populated. It is well known that high-resolution NWP forecasts can suffer from a double-penalty effect (e.g., Mittermaier et al., 2013), whereby a forecast of an event that is misplaced by a single grid box will count in a traditional contingency table as a miss and a false alarm. At lower resolution, the observation and forecast could potentially be in the same grid box and would therefore count as a hit. While gridded neighbourhood verification methods could be applied to lightning forecasts and gridded lightning observations, this could not yield entirely satisfactory outcomes due to the extreme representativeness mismatch between lightning observations and the model grid resolution. Neighbourhood methods apply an increasing degree of smoothing, which means that for larger neighbourhoods the actual location of any feature of interest within a neighbourhood becomes harder to pinpoint. Given the hazardous nature of lightning and localized impacts, this location information is rather important. Being more rooted or anchored to a location is relevant to several applications, such as for example river catchments, but it also seems an important constraint for forecasting the occurrence of lightning, for example, over Lake Victoria.

While upscaling to a much coarser grid than the model grid is a valid thing to do, in this study the representativeness of the observations was modified instead. Gridded lightning observations were 'smeared' out spatially to reflect the fact that although the location of an observed lightning stroke falls within a specific grid box, the lightning itself may have travelled some distance in *any* direction. A Gaussian kernel smoother implemented using the fast Fourier transform (FFT) is applied (via the `image.smooth` function in R; http://r-project.org) with a smoothing radius of 3 grid squares (0.12°) to represent the fact that lightning can travel ~10 km or more (e.g., Fuelberg et al., 2014). In terms of location accuracy, Zhu et al. (2017) found the ENGLN data to be accurate to 215 m based on a study over Florida. No statistics on location accuracy are available over Lake Victoria. This encompasses any error tolerances in the location detection as well. Taking this action reduces the spatial representativity differences between what the model diagnostic is providing and what the observation represents. Figure 3 provides a visual guide to the processing. In (a) an example observed 'maximum-in-the-day' lightning field is plotted with lightning strokes aggregated to the 4.4 km model grid, which shows the speckle nature of the field, even when diagnosing the maximum in the day at a location. To verify whether an event occurred, it has to be defined through the specification of a threshold. The simplest threshold is to verify the occurrence of 'any' lightning. The field in Figure 3a is converted to a binary field by applying a threshold of >0, that is 'any' lightning (not shown). As the kernel returns a density field, the intensity of the original field would be reflected in the smoothed one. To ensure that the density is consistent across multiple days (or forecast hours), smoothing is applied to the binary field. The outcome of applying the Gaussian kernel smoothing to the binary field is shown in (b). Using the binary field means that densities will always be constrained to less than 1 and makes the setting of the threshold for returning the smoothed field back to a binary one easier, as shown in (c). For the maximum-in-the-day fields, a value of 0.4 was used (after some pragmatic testing, comparing the size of the binary exceedance regions to those obtained by applying a threshold to the model fields). For comparison, the model maximum-in-the-day lightning flash density field is shown in (d) with the associated binary field in (e), with the same >0 threshold applied. Overall, this provides spatial structures on scales, which are a much better match to what the model is able to represent. These fields are used for computing all verification scores. To perform the verification (d) is compared with (f), the binary model field created after the threshold is applied.

3.2 | Scores

In this paper, a classical contingency table approach is used, but on the grid, so that the local characteristics of the forecast are preserved, and local biases and skill can be tracked. The hits (a), false alarms (b), misses (c) and correct rejections (d) are counted based on whether 'any' lightning was observed or forecast. For this study, a trio of metrics are reported: the frequency bias ($\mathrm{FB} = \frac{a+b}{a+c}$), the probability of detection ($\mathrm{POD} = \frac{a}{a+c}$) and the symmetric extremal dependence index (SEDI), defined as:

$$\mathrm{SEDI} = \frac{\log(\mathrm{POFD}) - \log(\mathrm{POD}) - \log(1-\mathrm{POFD}) + \log(1-\mathrm{POD})}{\log(\mathrm{POFD}) + \log(\mathrm{POD}) + \log(1-\mathrm{POFD}) + \log(1-\mathrm{POD})}, \quad (4)$$

where $\mathrm{POFD} = \frac{b}{b+d}$ is the probability of false detection. If there are few false alarms, POFD is very low, which is considered good. Ideally POFD is 0. The SEDI has many attractive attributes; it uses all the elements of the

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

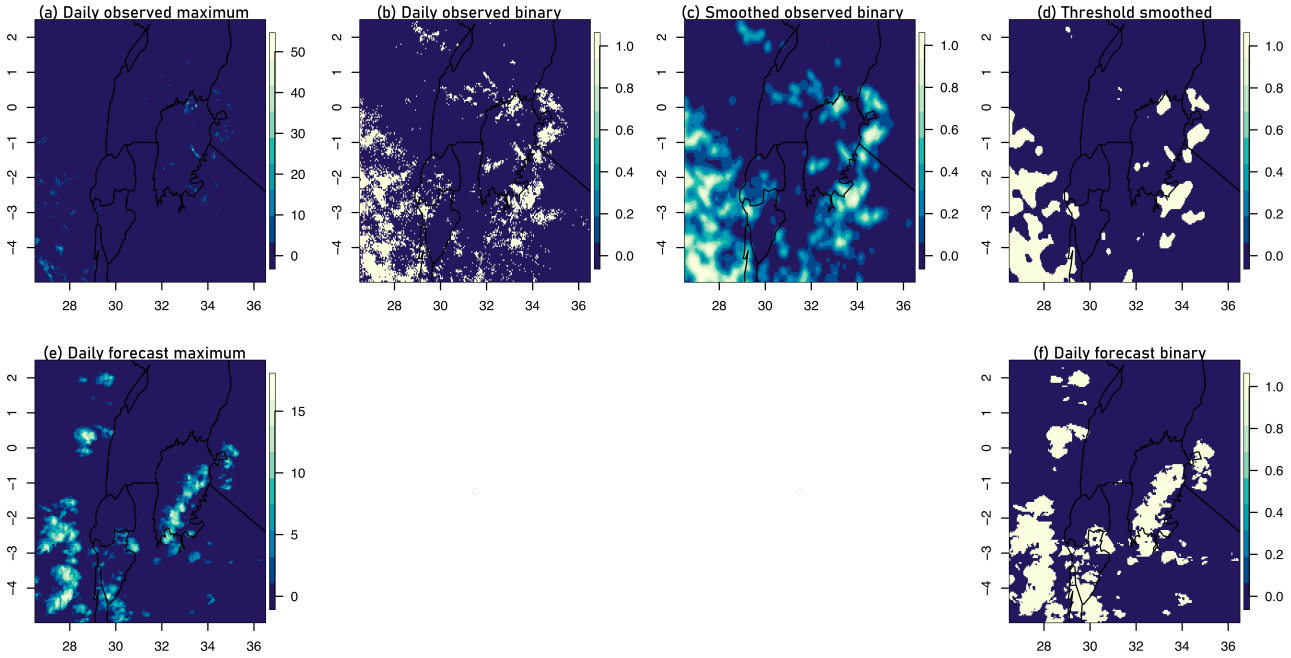

FIGURE 3 Illustrating the Gaussian kernel smoothing using the daily maximum-in-the-day gridded lightning fields. (a) the daily maximum lightning stroke density field on the 4.4 km model grid. Applying a >0 (any) threshold to (a) yield (b), which shows the speckle nature of the gridded lightning observation field, even when the strokes are aggregated to the model grid. (c) Applying a fast Fourier transform (FFT) smoothing of 3 grid squares (0.12°) to the field in (b). A threshold is applied to this smoothed field to return to a binary one, as shown in (d). For comparison, the model maximum daily lightning field is shown in (e) with the associated binary field in (f), with the same >0 (any lightning) threshold applied

contingency table, thus offering the juxtaposition of POD and POFD. The overall score will therefore depend on the tension between these two metrics. SEDI has a range of $[-1, 1]$. It is maximized when $\mathrm{POD} \to 1$ and $\mathrm{POFD} \to 0$. This can only be achieved when there are no zero contingency table entries. The minimum SEDI occurs when $\mathrm{POD} \to 0$ and $\mathrm{POFD} \to 1$. A score of less than zero means the forecast is less skilful than a random one. It is hard to define what constitutes a 'good' SEDI, as it will depend strongly on the application and context, but values above 0.7 should be considered good, and values between 0.4 and 0.7 could be considered as fair. SEDI is designed for evaluating rare (sparse) events, as it does not tend to zero when events are few; the 'any' lightning threshold used in this study can already be considered a rare event, in both space and time (due to the short duration and intermittency). Finally, SEDI is asymptotically equitable (i.e., all random forecasts will have the same score as the sample size tends to infinity). SEDI does have one weakness; in that it can become undefined if either POD or POFD or both are 0. This occurs when either a or b or both are 0 and, in a spatial context, will mean that SEDI cannot be computed for any grid point in the domain where this applies. For further information on SEDI, see Ferro and Stephenson (2011).

3.3 | The coverage–distance–intensity (CDI) method

Wilkinson (2017) introduced a new coverage–distance–intensity method, which is a threshold-based spatial method for evaluating model lightning forecasts from convective-scale models, finding that none of the existing spatial methods provided satisfactory results, since lightning density noting can vary over several orders of magnitude. Lightning forecasts are prone to large biases, which render many methods difficult to use. For example, when neighbourhoods are applied, the number of hits increases rapidly with increasing neighbourhood size, thus mitigating the double-penalty effect but at the expense of making the frequency bias a function of neighbourhood size. This may not be desirable. CDI treats each individual grid point separately while not adhering to the principle of exact grid-to-grid matching.

The CDI method somewhat defies the revised classification of spatial methods outlined by Dorninger et al. (2018) as it uses neither neighbourhoods or upscaling, nor does it define objects although it is analogous to the structure–amplitude–location (SAL) method introduced by Wernli et al. (2008). One of the components is essentially a distance metric. The reader is

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

referred to the original paper for more detail. Only the score formulations are provided here. Furthermore, the original score used to define coverage is replaced here by a new coverage score, which mirrors the formulation of the intensity score and avoids some of the issues that arise, especially when POD is zero, so that the trio of scores can be defined as follows:

$$C = \frac{P_m - P_o}{P_m + P_o}. \tag{5}$$

The coverage score C is based on the number of model and observed points P_m and P_o with lightning. If $C = 0$, the model and observations have the same number of grid squares with lightning; for $0 < C < 1$, the model is over-forecasting the areas with lightning, and for $-1 < C < 0$, the model is under-forecasting the areas with lightning. The intensity score I is based on the total number of lightning flashes in the model and observed T_m and T_o:

$$I = \frac{T_m - T_o}{T_m + T_o}. \tag{6}$$

Similar to C, if $I = 0$, the model and observations have the same number of lightning flashes; for $0 < I < 1$, the model is over-forecasting the number of lightning flashes, and for $-1 < I < 0$, the model is under-forecasting the number of lightning flashes. The third component, a quasi-symmetric distance skill score QSDS is defined as follows:

$$\mathrm{QSDS} = \begin{cases} 1, & D_{dis} = 0 \\ 1 - \frac{1}{2}\left[\frac{D_{dis}}{D_h} + \frac{ln(D_{dis})}{ln(D_h)}\right], & D_h > D_{dis} > 0 \\ 0, & D_{dis} = D_h \\ \frac{1}{2}\left[\frac{D_{dis} - D_m}{D_h - D_m} + \frac{ln(D_m - D_{dis})}{ln(D_m - D_h)}\right] - 1, & D_m > D_{dis} > D_h \\ -1, & D_{dis} = D_m \end{cases} \tag{7}$$

D_{dis} is the mean of the forecast-to-observed and observed-to-forecast displacement distance. D_h is the same for a completely hedged forecast (where lightning is forecast everywhere). If a model forecast has a D_{dis} value larger than D_h, it would be worse than forecasting lightning everywhere in the domain. The final quantity required for computing QSDS is D_m, which represents the largest possible separation distance between the model and observations and is taken as the mean distance of the observation locations to the four corners of the domain. Note that this implies that the domain is rectangular. QSDS = 1 indicates the forecast lightning is collocated with the observed lightning, while for QSDS ≤0, the location information has no skill. Note that for the CDI method, no Gaussian kernel dressing was applied to the lightning observations.

4 | RESULTS

4.1 | Comparing the diurnal cycle of lightning activity

It is illuminating to get an overall sense of how well the forecast diurnal cycle of lightning activity compares to the observed, as a bulk indicator of potential timing issues, and overall frequency of lightning. Given that there are distinct geographical modes of behaviour, it would also seem important to take this into consideration. The diurnal variations are explored in space and in time.

The accumulated spatial distribution of lightning is shown in Figure 4. These frequency maps were created by aggregating the total lightning over the 192 days between April and October 2019 into three 8-h windows: 06–14 UTC, 14–22 UTC and 22–06 UTC. Broadly speaking, the model (in panels (d)–(f)) captures the observed pattern (shown in panels (a)–(c)) fairly well with notably fewer lightning flashes. The period 14–22 UTC is associated with a minimum of lightning activity over the lake. The period 22–06 UTC is the most active in terms of lightning over the lake, whereas the regions to the NE, NW and W of the lake are fairly active irrespective of the time of day. It is noteworthy that the model has slightly higher activity over the lake between 06 and 14 UTC.

The diurnal domain average evolution of forecast and observed lightning flash density is provided in Figure 5, showing a number of different geographical and seasonal subsets. As stated earlier, the forecast itself is 54 h in length, initialized at 18 UTC, thus covering two full days. HIGHWAY and part 2 of this paper (see Mittermaier et al., 2021) are primarily interested in the lead times between $t + 13$ h and $t + 36$ h. This is the period that corresponds with the warnings that are issued, which are valid for 24 h from 06 UTC to 06 UTC. Although the focus of the verification statistics is also for the $t + 13$ h to $t + 36$ h period, it is interesting to see how the model behaves before and after this time to understand aspects of forecast conservation, for example.

The first row, panels (a)–(d) in Figure 5 shows the temporal evolution over a variety of subsets (the lake only, and the land areas to the west and east of the lake) but for the entire 192 days. Subsequent rows show the seasons separately: April–May (AM, June–July–August (JJA) and September–October (SO), noting that the two

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

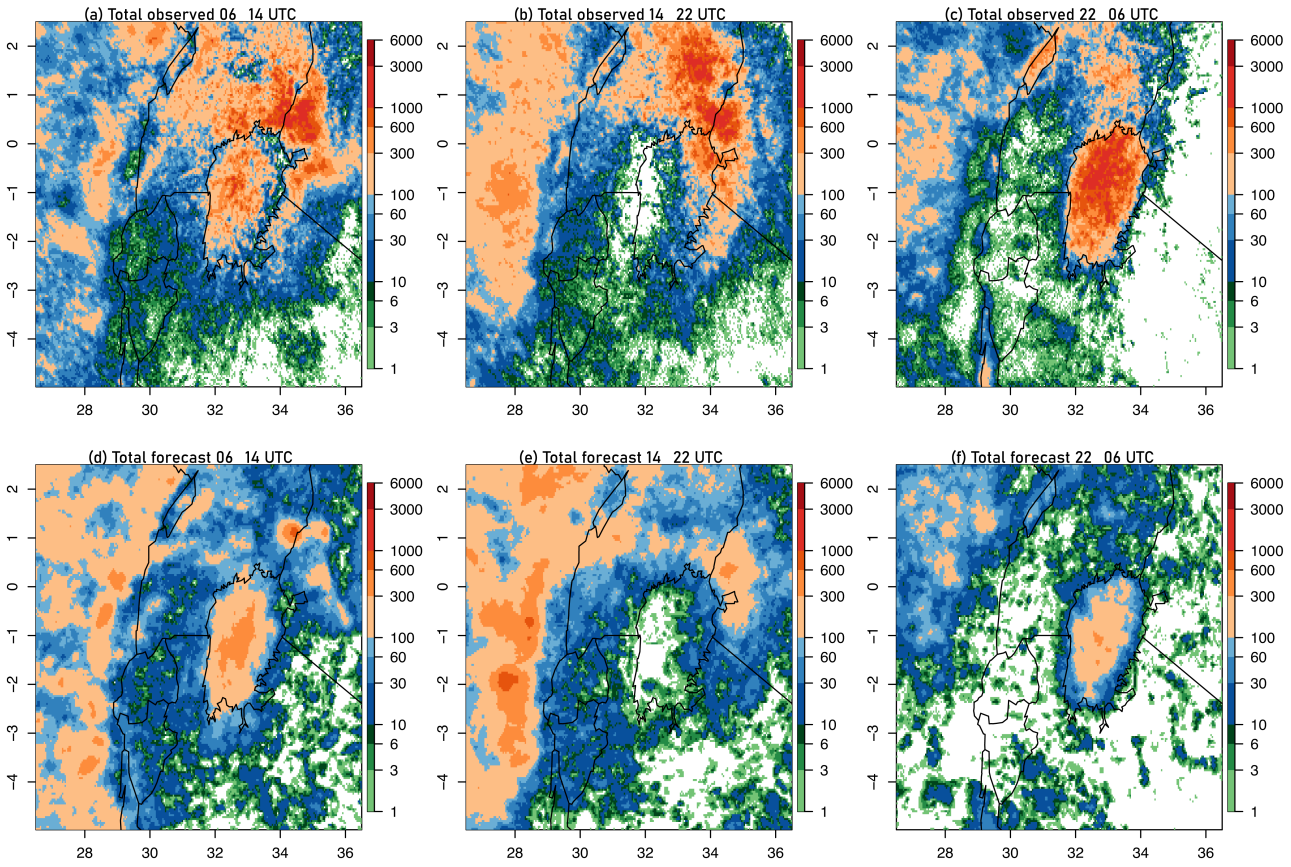

FIGURE 4 Total lightning flashes per model grid square for the period April–October 2019 split to reflect diurnal variations for EN observations in (a)–(c) and the forecast (d)–(f)

of the seasons are not complete. The geographical split was considered of interest given that Figures 2, 3 and 5 suggest that the observations show distinct differences in the two halves of the domain.

The main points from Figure 5 can be summarized as follows:

- The model lightning shows a distinct spin up in the first 6–12 h of the forecast (all panels). Thus, ignoring the first 6–12 h seems advisable.
- The domain-wide total lightning average (in (a), (e), (i) and (m)) is a juxtaposition of the different diurnal evolutions that exist over the lake and elsewhere and is quite difficult to interpret. The model does not capture this behaviour.
- Noting the differences in scaling for the lake panels, the observed peak in average lightning density over the lake is 3–8 times higher than elsewhere, depending on the season and location (comparing all panels).
- The model lightning domain average is generally lower than the observed domain average, with one notable exception in the region west of the lake.
- Overall, the average model forecast total lightning flash density is too high to the observed to the west of the lake (panels (c), (g), (k) and (o)). The timing of the peaks is fairly good, although the model lightning flash density tends to reduce too quickly in some seasons and overall.
- The model provides a reasonable guidance on the timing of the end of the diurnal peak where the forecast and observed total lightning domain averages are converging over the lake (panels (b), (f), (j) and (n)), but the peak is severely under-estimated.
- Over the lake, the forecast onset of the average peak is offset compared with observed average with some variations as a function of season: it is too late overall, slightly too early for AM (but fairly good agreement), too late for JJA and SO (panels (b), (f), (j) and (n)).
- The second forecast average peak in the 54 h forecast is always lower than the first (all panels).
- The shape of the observed average peak over the lake is markedly different from the one over land; while the observed total ramps up more slowly towards the peak

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

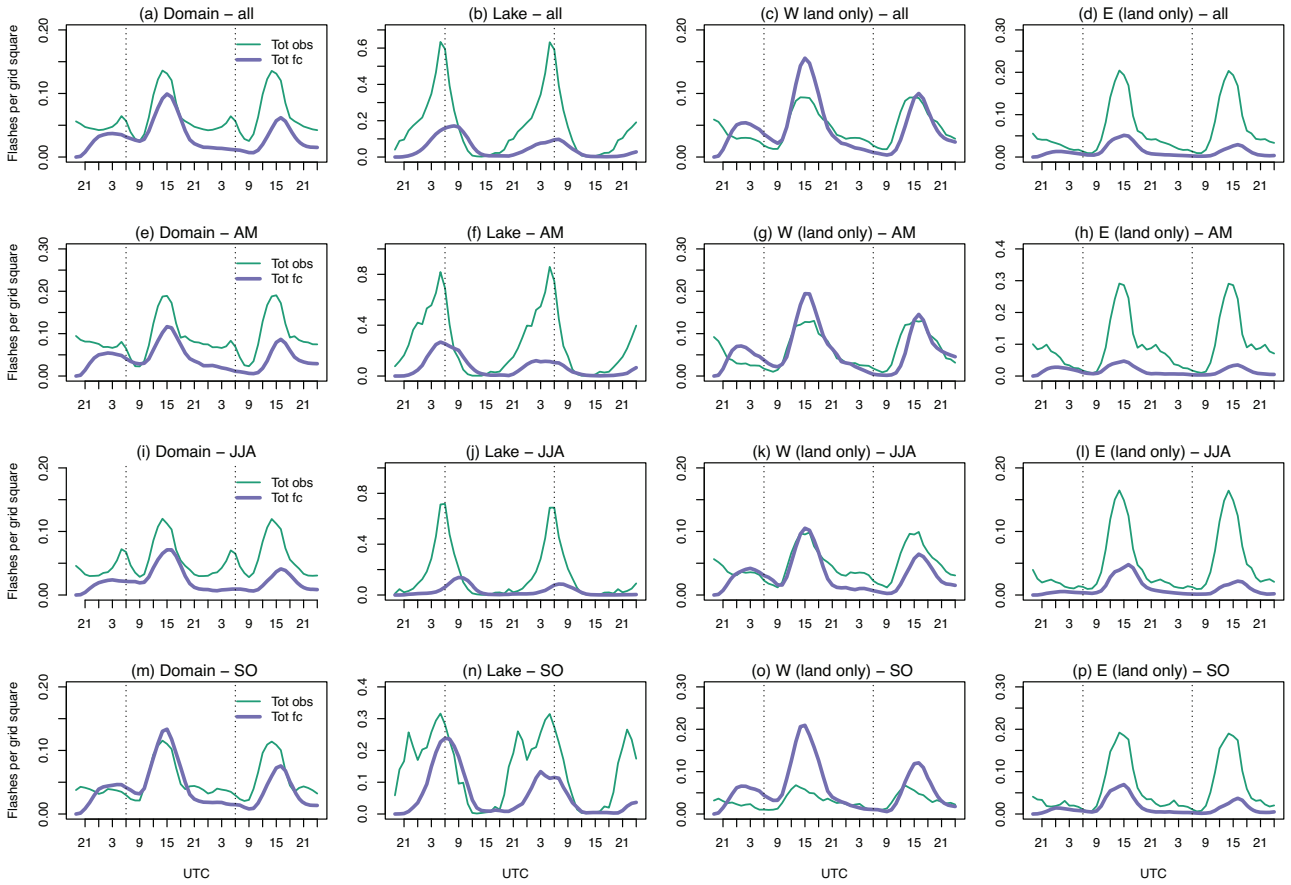

FIGURE 5 Diurnal cycle of lightning density partitioned over the domain. The whole study period (April–October) is shown as well as the seasons (AM, JJA and SO). Times are referenced with respect to the model run initialization time of 18 UTC. Results are provided for the whole domain, lake only and the land regions west and east of the lake (split at 31.48° E), given the large differences in orography). 'Day 1' is denoted by the vertical dotted lines on all the panels. Vertical dotted lines indicate the 'Day 1' boundaries 06 UTC ($t + 12$ h) and 06 UTC ($t + 36$ h) relevant for warnings

average and then drops off sharply over the lake (e.g., (b) and (f), the opposite is true for the land areas where there is a sharp rise and slow decline (e.g., (d) and (h)).

- For JJA, the forecast domain average peak is ∼4 h too late over the lake with notable timing offsets to the E of the lake as well (panels (j), (k) and (l)).
- The JJA observed total density domain average peak over the lake in (j) is narrowest with the sharpest increase (compared with (b) and (f) in particular).
- For SO, the observed total lightning domain average peak in (n) is broad and ragged over the lake. The forecast average fails to capture most of this peak. For the land areas to the west of the lake (o) suggests this is the only time where the forecast total density domain average truly appears to exceed the observed domain average, although again there is a timing offset in the peak.
- Away from the peaks, the model appears to keep more lightning going in both the full domain and the subregions.
- From the panels, it is clear that the definition of the warning period is not well synchronized with the diurnal evolution, especially for issuing warnings over the lake.

4.2 | Aggregated verification scores as a function of time of day

Figure 6 provides domain-wide aggregated hourly verification metrics for the presence of 'any' lightning. The results are shown for the entire period of April–October 2019 as well as for the separate seasons: AM, JJA and SO. Panels (a)–(c) show the results for the domain as a whole, whereas (d)–(f) show the results for the lake only. From the FB in (a), it is clear that over the domain as a

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

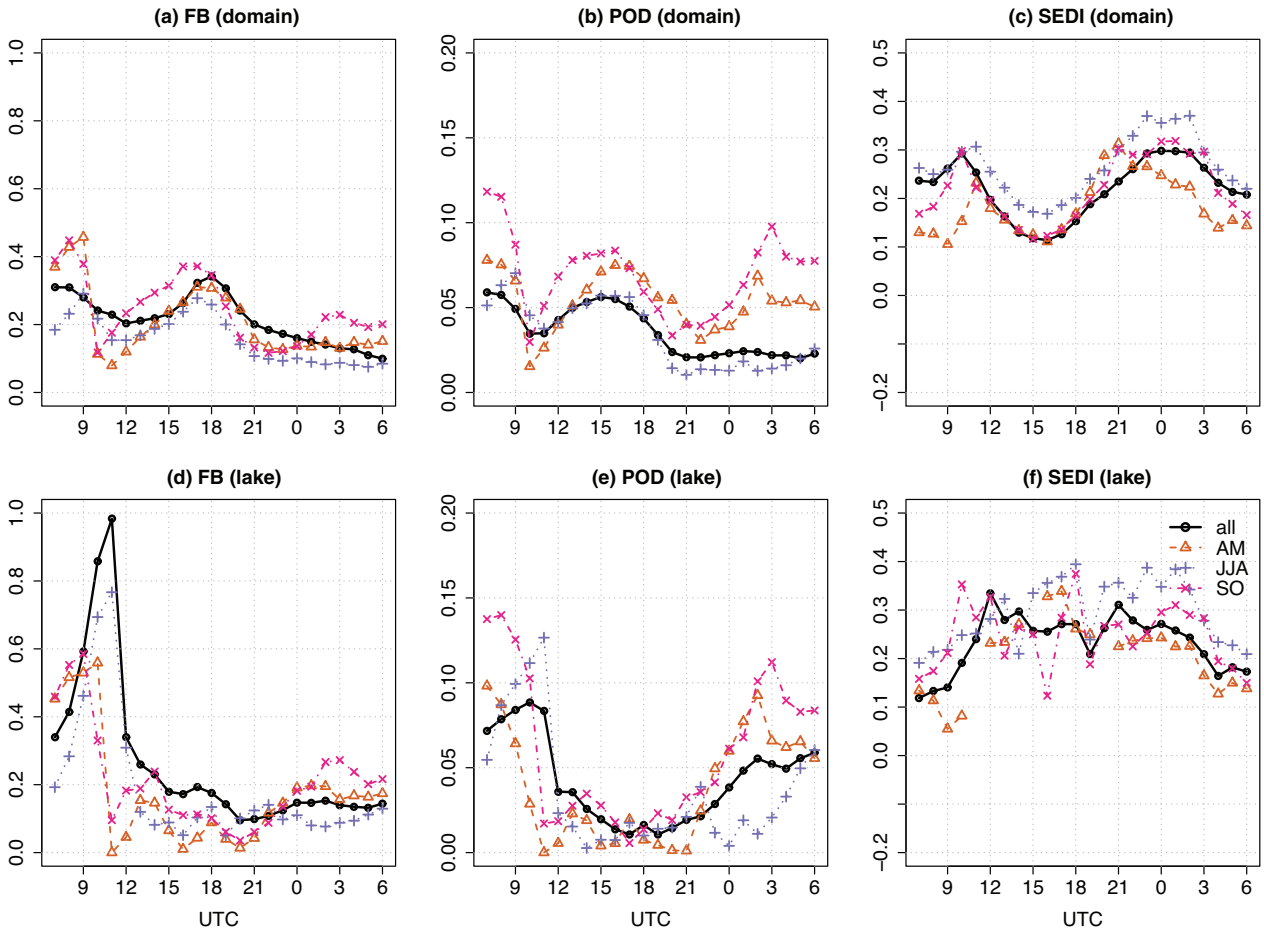

FIGURE 6 Diurnal evolution of spatially aggregated FB, POD and SEDI showing the results for the whole domain (a–c) and Lake Victoria only (d–f). Results are based on the individual hourly fields with varying lead times ($t + 13$ h to $t + 36$ h), all from the 18 UTC forecast initialization. Results are shown for the entire study period of 192 days between April and October, as well as for (partial) seasons: AM, JJA and SO

whole, there is too little lightning in the forecast with values between 0.2 and 0.4. There is also a decreasing trend with lead time, somewhat mitigated by 18UTC peak, which reflects the trends seen in Figure 4, that is the second forecast peak is lower than the first. Some seasonal variations exist, with SO having different timings to the other seasons and overall, again mirroring results in Figure 4. In (d), the FB over the lake is close to 1 (perfect) at 11 UTC, with a period of ∼3 h where the FB is above 0.5. False alarms are always less than misses (not shown), suggesting that the model is not producing enough locations with lightning, even over the lake. There is some seasonal variation in the peak of the FB, with AM and SO peaking somewhat earlier. The second diurnal peak in the forecast is far more diffuse in the FB values. Two measures of 'skill' are presented in Figure 6: POD and SEDI. Overall, the POD for the domain in (b) is very low, with values of ∼0.1 at best, peaking around mid- to late afternoon. The highest POD values are achieved for the SO season between 03 and 08 UTC. The low values generally demonstrate the difficulty of evaluating at the hourly resolution and at the grid-scale, even when the observations are suitably dressed (as they have been done here). Over the lake, the POD in (e) is somewhat higher with two distinct maxima at 10 UTC and towards 06 UTC the following day. The SEDI is more suited to rare or sparse events and shows better levels of skill overall, with values between 0.1 and 0.4 for the domain in (c). There is considerable spread in the hourly SEDI for the seasons, especially during the more active lightning periods overnight. In (f), the lake shows a different evolution in SEDI scores. Although the range of values is broadly similar, the SEDI values over the lake are generally lower for early morning hours, while they are considerably higher during the afternoon. From Equation (4), the SEDI is a combination of the POD and POFD. In this instance, the small POFD is enhancing the SEDI as the model is under-forecasting lightning generally, limiting

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

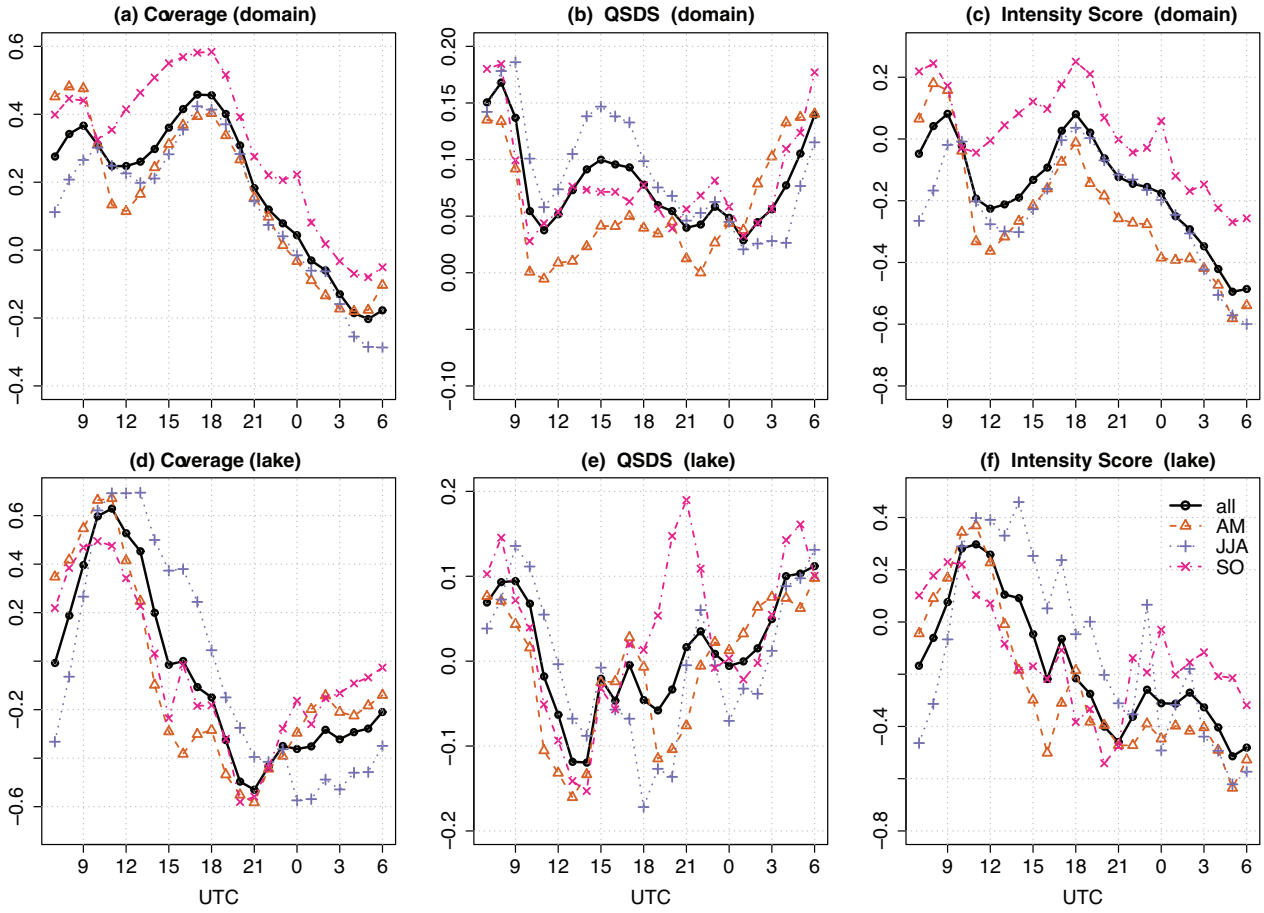

FIGURE 7 Same as for Figure 6 but for the component scores of the CDI method

the impact of false alarms, which would otherwise act to reduce SEDI.

4.3 | CDI components

The difficulties of strict matching of forecasts and observations in both space and time are well illustrated in Figure 6. This is why methods such as CDI are useful as they remove many of the constraints and provide more diagnostic information, separating out aspects of spatial coverage and intensity in particular. Figure 7 shows the CDI components for the domain as a whole (in panels (a)–(c)) as well as for the lake alone (panels (d)–(f)). Again, the results are shown for the entire period April–October 2019 and for the different seasons. The coverage scores C, that is, the area (grid squares) with lightning in the model is too large for substantial parts of the day. This is true for the domain overall and over the lake, as seen in (a) and (d). This is fairly consistent between the seasons and for the entire period. The model does, however, have too little lightning coverage between 00 and 06 UTC in (a) and between 18 and 06 UTC over the lake in (d). The lack of lightning coverage during hours of darkness is consistent with the findings from Hanley et al. (2020), which also mention the lack of rainfall at night, which they attribute to the lack of night-time storms. The results also support the finding that the lightning area reduces with lead time, since 06 UTC (t + 12 h) and 06 UTC (t + 36 h) are not matching up in magnitude (in an average sense). The lake coverage is much more marked in its peakedness (in terms of overforecasting coverage), which occurs at a different time to the domain as a whole and tends to have too large lightning areas between 07 and 15 UTC with much too little for the rest of the day. The peak in the coverage overforecasting of the area is at different hours, depending on the season, with JJA later and more prolonged than the others. JJA also has a more pronounced underforecasting of lightning coverage during the night and early morning hours. The QSDS is equal to 1 if the forecast and observed areas are collocated. In this instance, for the entire domain in (b), the values are, for the most part, greater than 0, suggesting that the forecasts are

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

spatially more skilful than forecasting lightning everywhere. Over the lake in (e), the results show negative QSDS values between 11 and 20 UTC, suggesting that during the day and early evening the location of forecast lightning over the lake (in relation to reality) is poor. The intensity scores plotted in (c) and (f) confirm what the FB showed in Figure 6 as well as the maps in Figure 5. For the domain as a whole, the intensity score is negative for most of the day, suggesting that the model lightning flash density is too low, where for a large portion of the day the model lightning density is spread over too large an area and it is too low. For the domain as a whole, only a few hours have close to the observed flash density: 09–12 UTC and 17–19 UTC. Over the lake, in (f), there is too much forecast lightning between 09 and 14 UTC. There are some seasonal variations, with SO showing a tendency towards over-forecasting of lightning intensity. Over the lake (f) shows that the model is over-forecasting intensity during daylight hours, with JJA showing an extension of this intensity over-forecast signal well into the evening. This is generally the period of minimum storm activity over the lake, suggesting the model is struggling to get the right level of activity at the right time. It also suggests that some of the timing issues need to be accounted for if the objective is to extract maximum skill (and value) from the forecast. More concerning perhaps is that during the most active overnight and early morning hours, the forecast lightning density over the lake is strongly under-estimated irrespective of season, which is at least partly due to the lack of coverage (storms).

4.4 | Spatial maps of categorical scores

Thus far, the categorical score results have followed the more conventional approach of representing aggregates of scores over a region for a given time. In this section, the spatial variations in skill and bias are explored, as these are highly relevant for a study like this, where the focus is very specific to a geographic location such as Lake Victoria. Bearing in mind sampling issues for individual grid points, squares that are white imply the score could not be computed.

In Figure 8, the diurnal cycle of the frequency bias (FB) in the hourly forecast fields (for all 192 forecasts in the study period) is plotted for the presence of 'any' lightning as the threshold (as before). For brevity, only every 3 h is shown. Note that the FB is plotted using a logarithmic colour scale because it can vary over several orders of magnitude. Regions that are blue and pale orange denote areas where the FB is near 1, that is where the observed and forecast frequency of occurrence are close to equal (ideal). While this suggests good correspondence over the eastern half of Lake Victoria at 09 UTC (12:00 LT), the forecast is way too active at 12 UTC (15:00 LT). The activity is confined to the western half of the lake at 15 UTC. Over the night and early morning hours, the FB is widely well below 1, suggesting significant under-forecasting, as shown in Figure 6.

The skill in the diurnal cycle is shown in Figure 9 using SEDI. Orange and red shades indicate high levels of skill with SEDI above 0.4. Moderate scores are shown in blue, with values in the 0.2–0.4 range. Large parts of the lake fall into this category at 09 UTC, with a cluster of higher scores at the southernmost tip of the lake. This region appears to be consistently associated with higher skill. Some pockets of higher skill persist over and around the western half and edge of the lake during the day. The mottled blue and green areas provide evidence of potential space (or more likely timing) errors, with very little over the lake until a return of convection and some skill from 00 UTC onwards, which progresses westwards with time.

Thus far, the analysis has focused on the characteristics of the hourly forecasts in the 24-h period between 06 and 06 UTC, made up of forecasts that are between $t + 13$ h and $t + 36$ h, which are subject to any timing errors as illustrated in Figure 5. As a final step, Figures 8 and 9 can be summarized by creating an overall aggregated score map from all the hourly forecast lead times that make up 'Day 1' forecasts together and aggregating over the entire study period but preserving the location information. What has also not been shown yet is the skill of a new forecast 'product', created by deriving the maximum-on-the-day (or 24 h period) for every grid square in the domain. This preserves the spatial information but removes the temporal dependence and focuses on the maximum that was forecast at any given location. This is shown in Figure 10a–c and can be contrasted to the hourly aggregated results in (d)–(f). To populate the contingency table, the maximum-in-the-day fields were still thresholded using the 'any' lightning (>0) threshold. Comparing the FB in (a) and (d), it is clear that using the maximum-in-the-day is a better match between the forecast and observed 'any' lightning with large parts with an FB near 1. Some pockets of over-forecasting exist. The lack of storms in the SE corner is also striking. There also appears some under-forecasting immediately to the east of the lake. As has been shown in previous sections, the FB in (d) shows a considerable amount of under-forecasting everywhere, even when using the lowest threshold possible. The POD in (b) shows the relatively high values over the lake, to the west and northeast, with isolated cases of storms

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

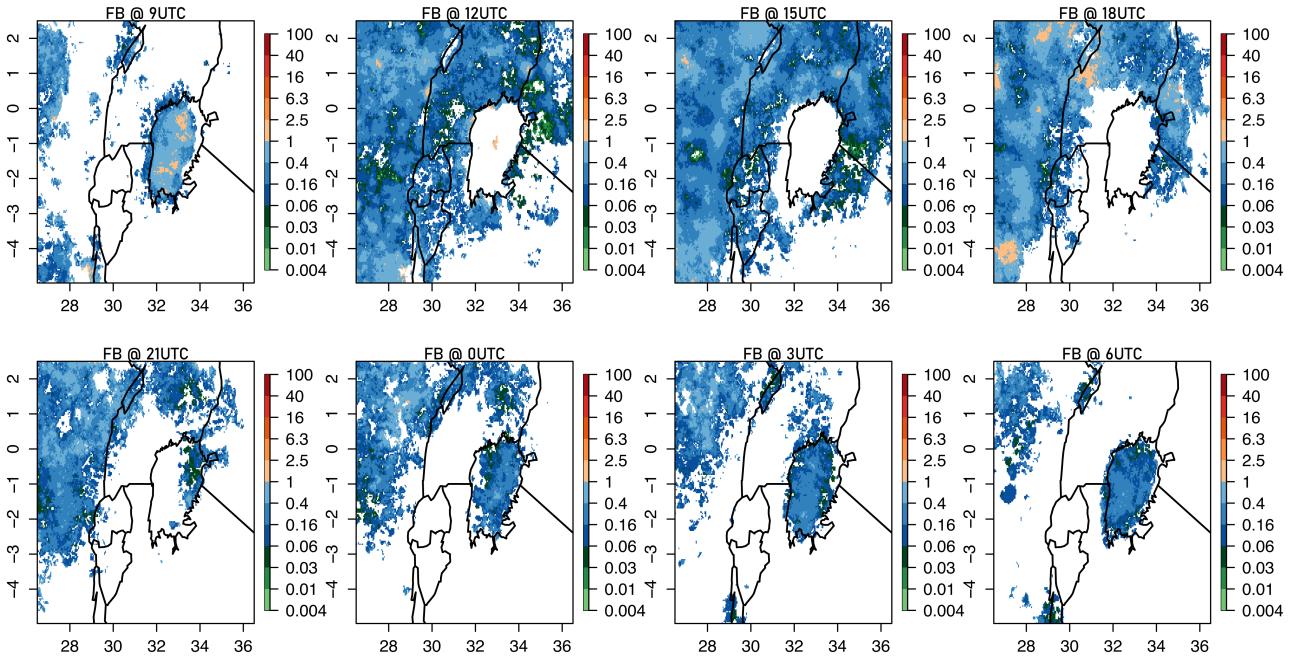

FIGURE 8 Diurnal evolution of the hourly frequency bias (FB) for the >0 threshold, that is any lightning for the period April–October 2019. White regions indicate grid points where the denominator ($a + c$) is 0 so that FB cannot be computed

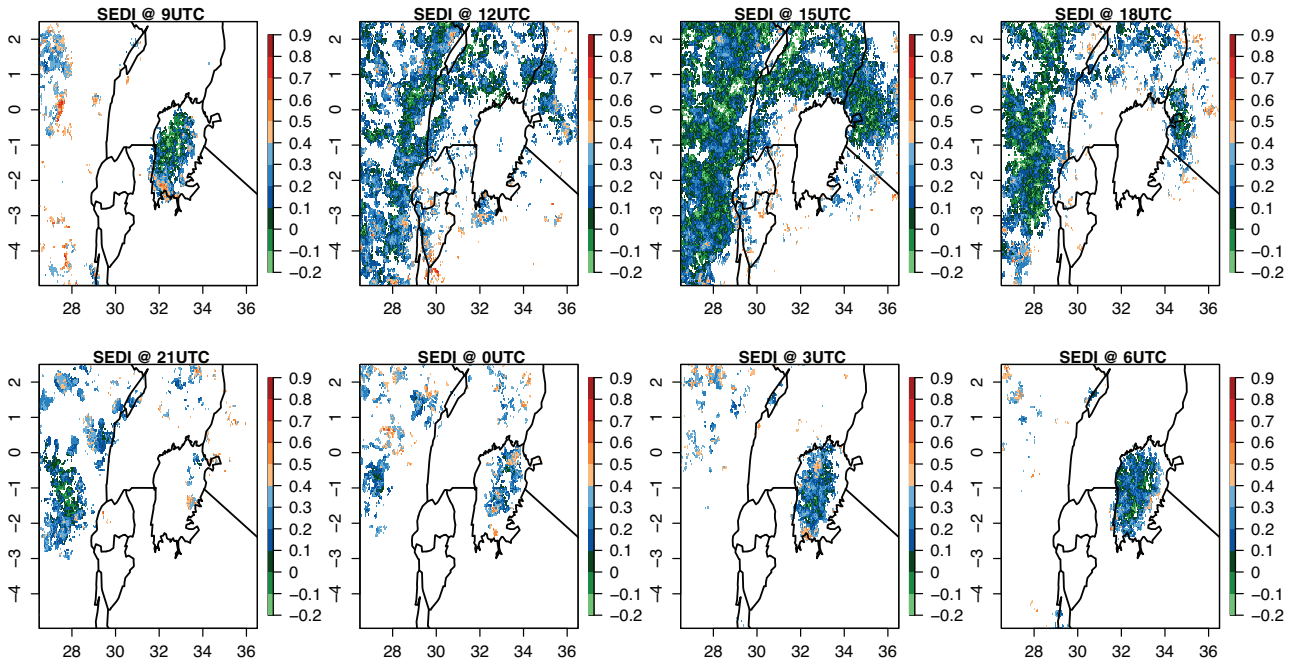

FIGURE 9 Diurnal evolution of the hourly SEDI for the >0 threshold (i.e., any lightning) for the period April–October 2019. White regions indicate grid points where either POFD or POD or both are zero so that SEDI cannot be computed

detected in the south and southwest. The hourly POD in (e) picks out the lake and the regions to the west and northeast but with much lower values as seen in Figure 5 also. Similarly, the SEDI in (c), which also accounts for the POFD, shows that the skill of the maximum-in-the-day field is much higher than the aggregated hourly values in (f), but there are some interesting differences in the distribution of scores. The

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

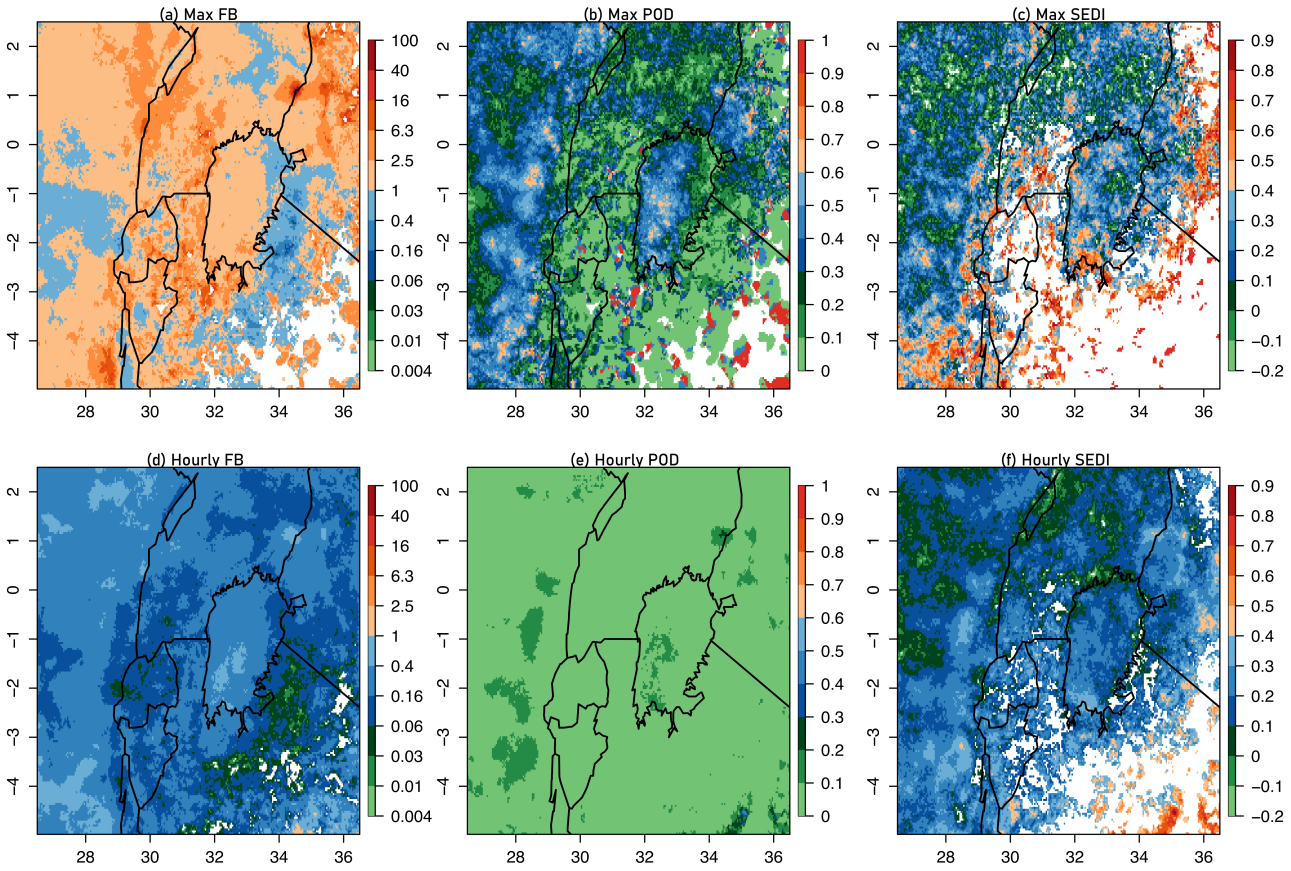

FIGURE 10 Spatial distributions of the frequency bias (FB), probability of detection (POD) and symmetric extremal dependency index (SEDI) for the April–October 2019 period. (a)–(c) show the result based on using the maximum-in-the-day for each grid square and (d)–(f) show the aggregated results for using the individual hourly fields with varying lead times ($t + 13$ h to $t + 36$ h)

SEDI in (c) shows higher scores, and the pattern is much sharper than that shown in (f). The average SEDI over the lake grid points in (c) is 0.19, whereas the average in (f) is 0.29, suggesting that timing errors are damaging to forecast skill but can be mitigated against by using a maximum-in-the-day representation. Of course, once the FB is addressed for the hourly fields, this may no longer be the case. The benefits of removing the time dependence are shown in Figure 10, showing how maximum skill can be extracted from the model forecasts, especially for the purposes of assisting in the issuing of warnings.

5 | CONCLUSIONS

This study presented some of the observed and forecast characteristics of lightning in the Lake Victoria region, which experiences some of the most intense storms in the world. Thus far, the UM lightning diagnostic has only been evaluated over the United Kingdom. This study presented a significant opportunity for understanding the model behaviour with a research-quality observation data set.

It has been shown that lightning observations are sparse (e.g., Figure 3, which shows a single hourly field), even when they are aggregated onto a grid, and this despite the region of interest being one of the most active for lightning in the world. Forecasting storms and lightning is challenging. For verification, the threshold was set to 'any' lightning to avoid issues with the clear intensity biases that were evident from the frequency maps.

The observed domain average lightning density shows that there are distinct geographical modes, which manifest themselves by having peaks in lightning activity at different times of the day, and these can also vary by season. The model peak is offset with respect to observed, especially over the lake, where the peak is too late by as much as 3–4 h. Overall, the model does not capture these different modes, with the model diurnal evolution very similar over the lake, and to the west and east. Overall, the model does not produce enough lightning, especially over Lake Victoria. There are a few exceptions, for example, to the west of the lake.

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

A classical contingency-table based categorical analysis of skill and bias was complemented with a newer spatial hybrid method, which compensates for the spatial and temporal mismatches that exist when verifying convective-scale models, where the detail looks realistic but is rarely at the right place at the right time. Given the prevalence and relative persistence of lightning in the Lake Victoria region, a comparison of 'old' and 'new' methods seemed appropriate as preserving the location information was key for this study. The need to preserve the location information and to constrain the frequency bias excluded commonly used neighbourhood methods where such information is not preserved. For the categorical analysis, the characteristics of the observations (sparsity and horizontal extent of lightning) are addressed by applying a Gaussian kernel smoothing to create a gridded field that (a) accounts for the spatial properties of lightning better and (b) is closer to what the forecast provides, which is a grid box average, thus reducing some of the extreme representativeness mismatches that exist (see, e.g., Marsigli et al., 2021). The CDI method did not use these dressed observation fields, although the original CDI method presented by Wilkinson (2017) was also enhanced through the introduction of a new coverage component score. The method was also applied to subdomains for the first time, as scores were computed for the domain as a whole as well as for only the lake itself. The CDI results suggest that the model lightning coverage is over-estimated while the intensity is under-estimated, especially at night.

When considering the hourly forecast fields, double-penalty errors are present, but these may have a temporal rather than just a spatial origin. They may be contributing to the low hourly frequency bias because this bias goes away when considering the 'maximum-in-the-day'. This may also be due to the lack of forecast lightning, which is clear with frequencies biases much below 1. The forecasts show some skill over the lake, to the west and northeast of the lake. Storm development over the lake is predictable. The strong clustering of scores in space are strong indicators of the preferred locations where storms occur, and these are very clearly delineated.

To conclude, there is useful skill in the maximum-in-the-day (24 h) model lightning density fields obtained from the hourly model output, and as demonstrated in the companion paper, these together with other model output can be beneficial for the issuing of warnings.

PERMISSION TO REPRODUCE MATERIAL FROM OTHER SOURCES

None required.

CONFLICT OF INTEREST

None.

AUTHOR CONTRIBUTIONS

Marion Mittermaier: Conceptualization (lead); formal analysis (lead); funding acquisition (lead); investigation (lead); methodology (lead); software (equal); supervision (lead); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Jonathan Wilkinson:** Conceptualization (supporting); formal analysis (supporting); investigation (supporting); methodology (equal); software (equal); validation (supporting); writing – original draft (equal); writing – review and editing (supporting). **Gabriella Csima:** Data curation (equal); formal analysis (supporting); investigation (supporting); software (equal); visualization (supporting). **Katrina Virts:** Data curation (equal); investigation (supporting); software (equal); visualization (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Steven Goodman:** Data curation (supporting); investigation (supporting); writing – original draft (supporting).

DATA AVAILABILITY STATEMENT

Forecasts are archived on internal Met Office storage systems and are subject to standard Met Office retention policy. Earth Networks and Vaisala were made available to project participants by the NASA Global Hydrology Resource Center Distributed Active Archive Center (GHRC DAAC). Data sets used in this study are available from the Met Office.

ORCID

Marion Mittermaier https://orcid.org/0000-0003-4752-3135
Jonathan Wilkinson https://orcid.org/0000-0002-6906-4999
Steven Goodman https://orcid.org/0000-0002-3315-9320
Katrina Virts https://orcid.org/0000-0002-8616-9557

REFERENCES

Albrecht, R.I., Goodman, S.J., Buechler, D.E., Blakeslee, R.J. & Christian, H.J. (2016) Where are the lightning hotspots on earth. *Bulletin of the American Meteorological Society*. 97(11), 2051–2068. https://doi.org/10.1175/BAMS-D-14-00193.1

Anyah, R.O., Semazzi, F.H.M. & Xie, L. (2006) Simulated physical mechanisms associated with climate variability over Lake Victoria basin in East Africa. *Monthly Weather Review*, 134, 3588–3609.

Ba, M.B. & Nicholson, S.E. (1998) Analysis of convective activity and its relationship to the rainfall over the rift valley lakes of East Africa during 1983–90 using the METEOSAT infrared channel. *Journal of Applied Meteorology*, 37, 1250–1264.

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Best, M.J., Pryor, M., Clark, D.B., Rooney, G.G., Essery, R.L.H., Mènard, C.B. et al. (2011) The joint UK land environment simulator (JULES), model description—part 1: energy and water fluxes. *Geoscientific Model Development*, 4, 677–699.

Boutle, I.A., Eyre, J.E.J. & Lock, A.P. (2014) Seamless stratocumulus simulation across the turbulent gray zone. *Monthly Weather Review*, 142, 1655–1668. https://doi.org/10.1175/MWR-D-13-00229.1

Bush, M., Allen, T., Bain, C., Boutle, I., Edwards, J., Finnenkoetter, A. et al. (2019) The first Met Office Unified Model/JULES regional atmosphere and land configuration, RAL1. *Geoscientific Model Development Discussion*, 2019, 1–47 Available at: https://www.geosci-model-dev-discuss.net/gmd-2019-130 [Accessed March 15, 2020]

Cecil, D.J., Goodman, S.J., Boccippio, D.J., Zipser, E.J. & Nesbitt, S. W. (2005) Three years of TRMM precipitation features. Part I: radar, radiometric, and lightning characteristics. *Monthly Weather Review*, 133(3), 543–566. https://doi.org/10.1175/MWR-2876.1

Chamberlain, J.M., Bain, C.L., Boyd, D.F.A., McCourt, K., Butcher, T. & Palmer, S. (2014) Forecasting storms over Lake Victoria using a high-resolution model. *Meteorological Applications*, 21, 419–430. https://doi.org/10.1002/met.1403

Dafis, S., Fierro, A., Giannaros, T.M., Kotroni, V., Lagouvardos, K. & Mansell, E. (2018) Performance evaluation of an explicit lightning forecasting system. *Journal of Geophysical Research—Atmospheres*, 123, 5130–5148. https://doi.org/10.1029/2017JD027930

Davies, T., Cullen, M., Malcolm, A., Mawson, M., Staniforth, A., White, A. et al. (2005) A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal Royal Meteorological Society*, 131, 1759–1782.

Dorninger, M., Gilleland, E., Casati, B., Mittermaier, M.P., Ebert, E. E., Brown, B.G. et al. (2018) The set-up of the mesoscale verification inter-comparison over complex terrain project. *Bulletin of the American Meteorological Society*, 99(9), 1887–1906.

Ebert, E.E. (2008) Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications*, 15, 51–64. https://doi.org/10.1002/met.25

Emersic, C. & Saunders, C.P.R. (2010) Further laboratory investigations into the relative diffusional growth rate theory of thunderstorm electrification. *Atmospheric Research*, 98, 327–340. https://doi.org/10.1016/j.atmosres.2010.07.011

Federico, S., Avolio, E., Petracca, M., Panegrossi, G., Sanò, P., Casella, D. et al. (2014) Simulating lightning into the RAMS model: implementation and preliminary results. *Natural Hazards and Earth System Sciences*, 14, 2933–2950. https://doi.org/10.5194/nhess-14-2933-2014

Ferro, A.T. & Stephenson, D.B. (2011) Extremal dependence indices: improved verification measures for deterministic forecasts of rare binary events. *Weather and Forecasting*, 26, 699–713.

Fiedler, E.K., Martin, M.J. & Roberts-Jones, J. (2014) An operational analysis of lake surface water temperature. *Tellus*, 66, 21247. https://doi.org/10.3402/tellusav66.21247

Fuelberg, H.E., Walsh, R.J. & Preston, A.D. (2014) The extension of lightning flashes from thunderstorms near Cape Canaveral, Florida. *Journal of Geophysical Research—Atmospheres*, 119, 9965–9979. https://doi.org/10.1002/2014JD022105

Giannaros, T.M., Lagouvardos, K. & Kotroni, V. (2017) Performance evaluation of an operational lightning forecasting system in Europe. *Natural Hazards*, 85, 1–18. https://doi.org/10.1007/s11069-016-2555-y

Gregory, D. & Rowntree, P. (1990) A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Monthly Weather Review*, 118, 1483–1506.

Griffiths, J. (Ed.). (1972) World survey of climatology. In: *Climates of Africa. Chapter 9—Eastern Africa*, Vol. 10. Amsterdam: Elsevier, pp. 313–332.

Groenemeijer, P., van der Velde, O., Tuschy, H., Gatzen, C., Dahl, J. & Verge, N. (2007) Verification of dichotomous lighting forecasts at the European Storm Forecasting Experiment (ESTOFEX). In: *4th European Conference on Severe Storms*, 10–14 September 2007, Trieste, Italy. Available at: https://www.researchgate.net/publication/266012686 [Accessed xxxx].

Hanley, K.E., Pirre, J.S.R., Bain, C.L., Hartley, A., Lean, H.W., Webster, S. et al. (2021) Assessment of convection-permitting versions of the Unified Model over the Lake Victoria basin region. *Quarterly Journal of the Royal Meteorological Society*, 147(736), 1642–1660.

Holle, R.L. & Murphy, M.J. (2017) Lightning over three large 468 tropical lakes and the strait of Malacca: exploratory analyses. *Monthly Weather Review*, 145, 4559–4573. https://doi.org/10.1175/MWR-D-17-0010.1

Huang, X.-Y., Barker, D., Webster, S., Dipankar, A., Lock, A., Mittermaier, M. et al. (2019) SINGV—the convective-scale numerical weather prediction system for Singapore. *ASEAN Journal on Science & Technology for Development*, 36(3), 81–90. https://doi.org/10.29037/ajstd.581

Lafore, J.P., Chapelon, N., Beucher, F., Kane, M.D., Gaymard, A., Kasimou, A. et al. (2017) West African synthetic analysis and forecast. In: Parker, D. & Diop-Kane, M. (Eds.) *Meteorology of tropical West Africa: the forecasters' handbook*. Wiley-Blackwell, pp. 423–451.

Lock, A., Brown, A., Bush, M., Martin, G. & Smith, R. (2000) A new boundary layer mixing scheme. Part 1: scheme description and single column model tests. *Monthly Weather Review*, 128, 3187–3199.

Lynn, B.H. (2017) The usefulness and economic value of total lightning forecasts made with a dynamic lightning scheme coupled with lightning data assimilation. *Weather and Forecasting*, 32(2), 645–663.

Lynn, B.H., Yair, Y., Price, C., Kelman, G. & Clark, A.J. (2012) Predicting cloud-to-ground and intracloud lightning in weather forecast models. *Weather and Forecasting*, 27, 1470–1488. https://doi.org/10.1175/WAF-D-11-00144.1

Lynn, B.H., Kelman, G. & Ellrod, G. (2015) An evaluation of using observed lightning to improve convective lightning forecasts. *Weather and Forecasting*, 30, 405–423. https://doi.org/10.1175/WAF-D-13-00028.1

Marsigli, C., Ebert, E., Ashrit, R., Casati, B., Chen, J., Coelho, C.A. S. et al. (2021) Observations of high-impact weather and their use in verification. *Natural Hazards and Earth System Sciences*, 21, 1297–1312. https://doi.org/10.5194/nhess-2020-362

McCaul, E.W., Goodman, S.J., LaCasse, K.M. & Cecil, D.J. (2009) Forecasting lightning threat using cloud-resolving model simulations. *Weather and Forecasting*, 24, 709–729. https://doi.org/10.1175/2008WAF2222152.1

McCaul, E.W., Priftis, G., Case, J.L., Chronis, T., Gatlin, P.N., Goodman, S.J. et al. (2020) Sensitivities of the WRF lightning

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

forecasting algorithm to parameterized microphysics and boundary layer schemes. *Weather and Forecasting*, 35, 1545–1560. https://doi.org/10.1175/WAF-D-19-0101.1

Mittermaier, M.P. (2008) The potential impact of using persistence as a reference forecast on perceived forecast skill. *Weather and Forecasting*, 23, 1022–1031.

Mittermaier, M.P., Roberts, N. & Thompson, S.A. (2013) A long-term assessment of precipitation forecast skill using the fractions skill score. *Meteorological Applications*, 20, 176–186.

Mittermaier, M.P., Landman, S., Csima, G. & Goodman, S. (2021) Convection-permitting numerical weather prediction and warnings over Lake Victoria. Part II: can model output support severe weather warning decision-making? *Meteorological Applications* [In press].

Mohan, G.M., Gayatri Vani, K., Hazra, A., Mallick, C., Chaudhari, H.S., Pokhrel, S. et al. (2021) Evaluating different lightning parameterization schemes to simulate lightning flash counts over Maharashtra, India. *Atmospheric Research*, 255, 105532. https://doi.org/10.1016/j.atmosres.2021.105532

Nag, A.M.J., Murphy, W.S. & Cummins, K.L. (2015) Lightning locating systems: insights on characteristics and validation techniques. *Earth and Space Science*, 2, 65–93. https://doi.org/10.1002/2014EA000051

Nicholson, S.E. (1996) A review of climate dynamics and climate variability in eastern Africa. In: Johnson, T.C. & Odada, E. (Eds.) *Limnology, climatology and paleoclimatology of the east African Lake*. Amsterdam: Gordon and Breach, pp. 57–78.

Ogallo, L. (1988) Relationship between seasonal rainfall in East Africa and the southern oscillation. *Journal of Climate*, 8, 31–43.

Prentice, S.A. & Mackerras, D. (1977) The ratio of cloud to cloud-ground lightning flashes in thunder-storms. *Journal of Applied Meteorology*, 16, 545–550. https://doi.org/10.1175/1520-450(1977)016<0545:TROCTC>2.0.CO;2

Qian, X. & Wang, H. (2021) Evaluation of different storm parameters as the proxies for gridded total lightning flash rates: a convection-allowing model study. *Atmosphere*, 12, 95. https://doi.org/10.3390/atmos12010095

Rakov, V. (2016) *Fundamentals of lightning*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139680370

Rawlins, F., Ballard, S., Bovis, K., Clayton, A., Li, D., Inverarity, G. et al. (2007) The Met Office global four-dimensional variational data assimilation scheme. *Quarterly Journal Royal Meteorological Society*, 133, 347–362.

Saunders, C.P.R., Bax-Norman, H., Emersic, C., Avila, E.E. & Castellano, N.E. (2007) Laboratory studies of the effect of cloud conditions on graupel/crystal charge transfer in thunderstorm electrification. *Quarterly Journal Royal Meteorological Society*, 132, 2653–2673. https://doi.org/10.1256/qj.05.218

Song, Y.I., Semazzi, F.H.M., Xie, L. & Ogallo, L.J. (2004) A coupled regional climate model for the Lake Victoria basin of East Africa. *International Journal of Climatology*, 24, 57–75.

Stratton, R.A., Senior, C.A., Vosper, S.B., Folwell, S.S., Boutle, I.A., Earnshaw, P.D. et al. (2018) A pan-African convection-permitting regional climate simulation with the Met Office Unified Model: CP4-Africa. *Journal of Climate*, 31, 3485–3508. https://doi.org/10.1175/JCLI-D-17-0503.1

Sun, X., Huang, X.Y., Gordon, C., Mittermaier, M., Beckett, R., Cheong, W.K. et al. (2020) A subjective and objective evaluation of model forecasts of Sumatra squall events. *Weather and Forecasting*, 35, 489–506. https://doi.org/10.1175/WAF-D-19-0187.1

Trewartha, G. (1961) *The Earth's problem climates*. Madison: University of Wisconsin Press, p. 334.

Virts, K.S. & Goodman, S.J. (2020) Prolific lightning and thunderstorm initiation over the Lake Victoria Basin in East Africa. *Monthly Weather Review*, 148, 1971–1985. https://doi.org/10.1175/MWR-D-19-0260.1

Virts, K.S., Wallace, J.M., Hutchins, M.L. & Holzworth, R.H. (2013) Highlights of a new ground-based, hourly global lightning climatology. *Bulletin of the American Meteorological Society*, 94, 1381–1391. https://doi.org/10.1175/BAMS-D-12-00082.1

Walters, D.N., Best, M.J., Bushell, A.C., Copsey, D., Edwards, J.M., Falloon, P.D. et al. (2011) The Met Office Unified Model global atmosphere 3.0/3.1 and JULES global land 3.0/3.1 configurations. *Geoscientific Model Development*, 4, 1213–1271.

Wernli, H., Paulat, M., Hagen, M. & Frei, C. (2008) SAL—a novel quality measure for the verification of quantitative precipitation forecasts. *Monthly Weather Review*, 136, 4470–4487.

Wilkinson, J.M. (2017) A technique for verification of convection-permitting NWP model deterministic forecasts of lightning activity. *Weather and Forecasting*, 32, 97–115. https://doi.org/10.1175/WAF-D-16-0106.1

Wilson, D. & Ballard, S. (1999) A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Quarterly Journal Royal Meteorological Society*, 125, 1607–1636.

Wilson, D.R., Bushell, A.C., Kerr-Munslow, A.M., Price, J.D. & Morcrette, C.J. (2008) PC2: a prognostic cloud fraction and condensation scheme. I: scheme description. *Quarterly Journal Royal Meteorological Society*, 134, 2093–2107.

Wood, N., Staniforth, A., White, A., Allen, T., Diamantakis, M., Gross, M. et al. (2014) An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quarterly Journal Royal Meteorological Society*, 140, 1505–1520.

Woodhams, B.J., Birch, C.E., Marsham, J.H., Bain, C.L., Roberts, N.M. & 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, 2757–2780. https://doi.org/10.1175/MWR-D-17-0396.1

Yin, X., Nicholson, S.E. & Ba, M. (2000) On the diurnal cycle of cloudiness over Lake Victoria and its influence on evaporation from the lake. *Hydrological Sciences Journal*, 45, 407–424.

Zhu, Y., Rakov, V.A., Tran, M.D., Stock, M.G., Heckman, S., Liu, C. et al. (2017) Evaluation of ENTLN performance characteristics based on the ground truth natural and rocket-triggered lightning data acquired in Florida. *Journal of Geophysical Research—Atmospheres*, 122, 9858–9866. https://doi.org/10.1002/2017JD027270

Zipser, E.J., Cecil, D.J., Liu, C., Nesbitt, S.W. & Yorty, D.P. (2006) Where are the MOST intense thunderstorms on earth? *Bulletin of the American Meteorological Society*, 87(8), 1057–1072.

How to cite this article: Mittermaier, M., Wilkinson, J., Csima, G., Goodman, S., & Virts, K. (2022). Convective-scale numerical weather prediction and warnings over Lake Victoria: Part I—Evaluating a lightning diagnostic. *Meteorological Applications*, *29*(3), e2038. https://doi.org/10.1002/met.2038

14698080, 2022, 3, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.2038 by CochraneChina, Wiley Online Library on [03/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License