Environmental Conditions Affecting Global Mesoscale Convective System Occurrence

MARK R. MUETZELFELDT,[a] ROBERT S. PLANT,[a] HANNAH M. CHRISTENSEN,[b] ZHIXIAO ZHANG,[b] TIM WOOLLINGS,[b] ZHE FENG,[c] AND PUXI LI[d,e]

[a] *Department of Meteorology, University of Reading, Reading, United Kingdom*
[b] *Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, United Kingdom*
[c] *Atmospheric, Climate and Earth Sciences Division, Pacific Northwest National Laboratory, Richland, Washington*
[d] *State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, China Meteorological Administration, Beijing, China*
[e] *LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China*

ABSTRACT: The ERA5 environments of mesoscale convective systems (MCSs), tracked from satellite observations, are assessed over a 20-yr period. The use of a large set of MCS tracks allows us to robustly test the sensitivity of the results to factors such as region, latitude, and diurnal cycle. We aim to provide novel information on environments of observed MCSs for assessments of global atmospheric models and to improve their ability to simulate MCSs. Statistical analysis of all tracked MCSs is performed in two complementary ways. First, we investigate the environments when an MCS has occurred at different spatial scales before and after MCS formation. Several environmental variables are found to show marked changes before MCS initiation, particularly over land. The vertically integrated moisture flux convergence shows a robust signal across different regions and when considering MCS initiation diurnal cycle. We also found spatial scale dependence of the environments between 200 and 500 km, providing new evidence of a natural length scale for use with MCS parameterization. In the second analysis, the likelihood of MCS occurrence for given environmental conditions is evaluated, by considering all environments and determining the probability of being in an MCS core or shield region. These are compared to analogous non-MCS environments, allowing discrimination between conditions suitable for MCS and non-MCS occurrence. Three environmental variables are found to be useful predictors of MCS occurrence: total column water vapor, midlevel relative humidity, and total column moisture flux convergence. Such relations could be used as trigger conditions for the parameterization of MCSs, thereby strengthening the dependence of the MCS scheme on the environment.

SIGNIFICANCE STATEMENT: Large storm systems called mesoscale convective systems form across Earth. These are collections of thunderstorms, with associated high-level clouds that produce substantial, lighter rainfall and modulate Earth's energy balance. They produce hazardous weather conditions, such as floods and high winds, and are responsible for a high percentage of rainfall in many regions globally. We investigate the environmental conditions under which they form, so that we can understand the spatial extent of the environment which is important for their formation, and also where and when the effects of these storms might be felt. The novel information generated here should help improve the representation of these storms in weather and climate models, improving the prediction of rainfall, thunderclouds, and high-level clouds.

1. Introduction

Mesoscale convective systems (MCSs) play a key role in the global hydrological cycle, providing over 50% of precipitation over large areas of the tropics (Nesbitt et al. 2006; Feng et al. 2021) and also providing substantial fractions at higher latitudes over locations including the U.S. Great Plains, southeast China, south Brazil, and Argentina (Fritsch et al. 1986; Rasmussen et al. 2016; Yun et al. 2021; Feng et al. 2021; Kukulies et al. 2023). They are often the cause of hazardous weather conditions such as intense precipitation extremes and wind bursts (e.g., Maddox 1980; Schumacher and Rasmussen 2020). Thus, understanding the environmental conditions under which they form could help to better understand the global precipitation distribution and the drivers of certain extreme weather events.

There are two reasons why the environmental conditions of MCSs are of interest for this work: to understand how the environment changes before and during MCSs' lifetime and to derive predictive relationships between the environment and

---

MCS occurrence. Both of these can help improve the representation of MCSs in atmospheric models. The first can be used to check that models are producing the correct conditions for MCS formation and over what spatial scales these conditions are important. The second can be used to predict when a model should be producing MCSs, based on the environmental conditions which are present.

Both motivating reasons may be important at a range of model grid spacings. MCSs form from the interaction of many scales of motion, from individual cumulus clouds to large collections of cumulonimbus clouds, and so representing this faithfully in models requires representing all of these scales (McAnelly and Cotton 1986; Houze 2004, 2018). At tens of kilometer grid spacing, the upscale organization is not properly resolved (Moncrieff and Liu 2006; Li et al. 2023a), even though this is much smaller than the size of an MCS, and thus, one might expect them to be resolved. If the finer scales are not represented, then the effects of MCSs will have to be parameterized. Since the current generation phase 6 of the Coupled Model Intercomparison Project (CMIP6) models have a finest grid spacing of around 25 km for long-term simulations (i.e., pushing the boundaries of what is currently computationally feasible) [e.g., High Resolution Model Intercomparison Project (HighResMIP); Haarsma et al. 2016], this will likely be the case for years to come in climate models. Furthermore, as high-resolution, convection-permitting models become available, it will still be necessary to run lower-resolution models for longer-duration or larger ensemble simulations, due to computational cost. MCS parameterization schemes exist (Moncrieff et al. 2017; Chen et al. 2021) but are not in widespread use. Other attempts have been made to model the organization of convection through a feedback between previous convection and the parameterized convective entrainment rate at the next time step (Mapes and Neale 2011; Park 2014; Daleu et al. 2023). Here, we aim to provide evidence of when both classes of schemes should be active (so-called "trigger" conditions) as well as what changes should be expected in the environment before and while simulated MCSs occur and the length scales over which the environment affects MCS formation.

Improving the representation of MCSs in models may lead to a range of improvements because MCSs can couple to their larger-scale environments in many ways. MCSs are primarily driven by deep convection, with a mesoscale circulation producing a large stratiform region that further organizes and strengthens the deep convection. The stratiform region is associated with a large release of latent heat in the upper troposphere and evaporative cooling below, leading to a characteristic "top-heavy" heating profile (Schumacher et al. 2004; Houze 2004), which varies with the amount of stratiform precipitation in both the tropics and midlatitudes (Liu et al. 2021). The top-heavy profile couples to dynamics in distinctive ways. On a local scale, it causes a gravity bore which has a slower phase speed than the corresponding bore from deep convection, potentially leading to the upscale transport of energy ("gregarious convection"; Mapes 1993). On a global scale, the top-heavy heating can impact the general circulation (Schumacher et al. 2004). MCSs also couple to shortwave and longwave radiation, creating a strong cooling effect at the top of the atmosphere (Feng et al. 2011; Wall et al. 2018). Furthermore, MCSs, particularly those that form under the influence of easterly waves such as the African easterly wave, can act as seeds for tropical cyclones (Gray 1998; Núñez Ocasio et al. 2020; Enyew and Mekonnen 2022). MCSs are also a major contributor to the diurnal cycle of convection (e.g., Mohr and Zipser 1996; Li et al. 2021), and so improving their representation in models may help alleviate the bias in the representation of the diurnal cycle present in many global models (Muetzelfeldt et al. 2021b). Understanding the environmental conditions that lead to upscale growth of convective clouds to larger MCSs can therefore serve the purpose of indicating where these effects will be felt.

Previous studies have found relationships between environmental conditions and MCS formation and maintenance. Chen et al. (2017) found that the total column water vapor (TCWV) is the strongest predictor of precipitation produced by large-scale precipitation features (roughly equivalent to MCSs), with relative humidity (RH) at low and midlevels and low-level shear also having predictive value. CAPE and convective inhibition (CIN) are both weak predictors in their analysis. Chen et al. (2023) considered vertical profiles of moisture, equivalent potential temperature θ_e, temperature, vertical ascent, and shear, finding that enhanced lower-level moisture, a warmer midtroposphere, stronger ascent, and stronger deep shear are all associated with higher MCS precipitation. Several other studies have shown links between TCWV and MCS or MCS precipitation (e.g., Peters et al. 2009; Ahmed and Schumacher 2015; Schiro et al. 2020; Chen et al. 2022). Galarneau et al. (2023) found relationships between precursor environments and MCSs over the tropics, including an increase in moisture and mesoscale ascent before MCS formation. Many studies have shown relationships between vertical wind shear and MCSs (e.g., LeMone et al. 1998; Chen et al. 2023), and the detailed structure of the vertical profile of wind (as viewed by, say, a hodograph) has also been shown to play a role (Muetzelfeldt et al. 2021a).

Studies of favorable environmental conditions for MCSs are often based on a given region. For example, Laing and Fritsch (2000) analyze the large-scale environments over five MCS population centers, Africa, Australia, China, South America, and the United States, finding that vertical wind shear and CAPE are important, and there is often a low-level jet of high θ_e air that undergoes forced ascent over a cold pool. Peters and Schumacher (2014) analyze MCSs over the central and eastern United States, finding that they could classify the MCSs into two types, warm season and synoptic, consistent with the previous findings over this region (e.g., Fritsch and Forbes 2001). Li et al. (2023b) analyzed MCSs over East Asia, finding that large-scale patterns in moisture are important for the distributions of MCSs over this region. Likewise, many case studies of individual MCSs have been done (e.g., Houze 1977; Schwendike et al. 2010). From these studies, we can draw lessons about which environmental variables might be important, but due to their limited geographical reach or sample size, we cannot draw robust conclusions about the global environments under which MCSs form.

From previous studies, we have identified a set of environmental variables that may be important for MCS occurrence,

covering both the thermodynamic and dynamic environments, including TCWV, CAPE, shear between different levels, and moisture flux convergence (see section 2b for a full list). Our work adds to previous studies on precursor environments by decomposing our results by diurnal cycle, season, latitude bands, and land–sea splits globally. Furthermore, in a similar way to Chen et al. (2017), we perform analysis conditional on the given environmental conditions. However, we do this to produce probabilistic likelihoods of MCS occurrence for given conditions, instead of probabilities of rainfall, as they do.

The aim of this study is therefore to improve our understanding of the environments under which MCSs form. This will lead to a better understanding of observed MCSs and should also make it possible to determine whether atmospheric models (both high resolution and coarser) are producing realistic precursor environmental conditions for MCS initiation. This analysis will be informative for determining what the conditions are when an MCS has formed (i.e., it is conditional on an MCS being present) but will not be useful for determining what the probability of MCS occurrence is for a given environmental condition. To address this, we perform analysis which inverts the question, asking what the likelihood of MCS occurrence is conditional on the environment. For both analyses, we perform decomposition by diurnal cycle, season, and region, to ascertain how general our results are. This allows us to propose a design for a probabilistic trigger condition for when an MCS parameterization scheme, such as the multiscale coherent parameterization scheme (MCSP; Moncrieff et al. 2017), should be active.

The rest of this paper is laid out as follows. Section 2 describes the MCS tracking dataset and the ERA5 variables that we have chosen to represent the environment, as well as how these are combined to derive the environmental conditions of MCSs. We then set out our main results, first focusing on the precursor analysis in section 3 before describing the probabilistic trigger results in section 4. We finish with a summary in section 5.

2. Methods

a. MCS tracking dataset

We use version 2 of the MCS dataset (Feng 2023) described in Feng et al. (2021), which uses the PyFLEXTRKR tracking algorithm to track deep convection (Feng et al. 2023a). The version 2 enhancement means that the analysis is performed globally, rather than in three separate regions and then combined, as in version 1 (Feng et al. 2021). The dataset covers the period from June 2000 to December 2020, and we use the complete years 2001–02 and 2006–20 for our analysis, sampling both interannual variability as well as both El Niño/La Niña conditions. As described in Feng et al. (2021), section 5c, we exclude the years 2003–05 due to frequently missing satellite brightness temperature data over East Asia. The spatial coverage of 60°S–60°N means that we can perform a near-global analysis of MCSs, including continental midlatitude MCS in the United States and South America.

The MCS dataset is created by first detecting potential MCS cloud features at each time using the NASA merged geostationary satellite brightness temperature T_b (Janowiak et al. 2023) and combining this with precipitation data from IMERG (Huffman et al. 2015, using the "final" version 06). Both datasets are used at hourly frequency, on the half hour (e.g., at 1230 UTC). The 4-km T_b dataset is upscaled to the 0.1° resolution of IMERG. A threshold of $T_b < 241$ K is used to define a cold cloud system (CCS) mask, which determines the size of any cloud system. Furthermore, an MCS is required to have a convective core, defined by $T_b < 225$ K. These thresholds were chosen to capture high-level clouds and deeper convection, respectively, and are the same as in a previous study (Feng et al. 2021). We make use of both of these criteria later to define an MCS shield, equivalent to the CCS but excluding the core, and an MCS core. We further define a non-MCS system using the same definitions, except that there are no requirements for system size and lifetime.

PyFLEXTRKR (Feng et al. 2023a) is used to track cloud objects from one time step to another. Within each CCS mask, precipitation features are defined as contiguous areas with rain rate > 2 mm h^{-1}. There are three key criteria that must be met for a tracked object to be classed as an MCS:

1) CCS area $> 4 \times 10^4$ km^2 with a precipitation feature $>$ 100 km on its major axis;
2) lifetime ≥ 4 h; and
3) lifetime-dependent thresholds on precipitation feature area, mean rain rate, rain rate skewness, and heavy rain volume ratio.

The lifetime-dependent thresholds are determined by calibration of the MCS tracking datasets from independently tracked MCSs in the U.S. Next Generation Weather Radar (NEXRAD) network (Bowman and Homeyer 2017) and are validated by using a similar radar network over China (Chen et al. 2020). The area threshold is relatively strict, which means that only the larger MCSs are tracked compared to other studies (Yuan and Houze 2010; Huang et al. 2018). Using larger systems suits our study because we wish to capture the large-scale environmental conditions associated with MCSs. The lifetime threshold of 4 h is in line with other studies. Because it is a tracking dataset, we can extract information on how the environment changes over the course of an MCS lifetime.

By comparing candidate MCS locations with the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al. 2010), tropical cyclone tracks are filtered out by excluding those that overlap with tropical cyclones. Similarly, land-falling atmospheric rivers in the west coast of North America, South America, and Europe are filtered out by comparing with a dataset containing atmospheric rivers (Rutz et al. 2019) and excluding those that overlap.

b. ERA5 environmental conditions

To sample environmental conditions surrounding MCSs, we use ERA5 (Hersbach et al. 2020), which is available for the 20 years of our study. Variables are available on a 0.25° grid, at hourly frequency (on the hour). The large size of the tracked MCS means that an MCS covers at least around 50

TABLE 1. ERA5-derived environmental variables and their definitions, which are used to investigate the environmental conditions close to MCSs. Here, ρ is the density, q is the specific humidity, $\mathbf{u}$ is the vector horizontal wind, RH is the relative humidity, and θ_e is the equivalent potential temperature. Vertical integrals are from the surface to the model top unless otherwise stated.

Name	Description	Definition
TCWV	Total column water vapor	Vertical integral of water vapor
CAPE	Convective available potential energy	CAPE of most unstable layer from surface to 350 hPa
CIN	Convective inhibition	As CAPE, but for negative buoyancy
MFC	Vertically integrated moisture flux convergence	Vertical integral of $\nabla\cdot(\rho q\mathbf{u})$
LLS	Low-level shear	Wind diff between surface and 800 hPa
M2HS	Mid-to-high shear	Wind diff between 600 and 400 hPa
DS	Deep shear	Wind diff between surface and 400 hPa
RHlow	RH at low levels	Average RH between 1000 and 850 hPa
RHmid	RH at midlevels	Average RH between 700 and 400 hPa
$\theta_e$mid	θ_e at midlevels	Average θ_e between 850 and 750 hPa
Δ 3-h TCWV	3-h change in TCWV	
Δ 3-h CAPE	3-h change in CAPE	

ERA5 grid cells, and so spatial averaging over an MCS should not be sensitive to small-scale fluctuations. In the vertical, 137 levels from the surface to the model top (at approximately 1 Pa) are used. We have used the standard ERA5 atmosphere to translate ERA5 model levels to pressure levels throughout. This is an excellent approximation over sea and remains a good approximation over land, although the pressure levels that correspond to a given model level are higher in altitude due to the presence of orography (see Table S1 in the online supplemental material for more information). Using model levels also removes the need to handle the intersection of a given pressure level with the surface.

The full list of environmental variables is set out in Table 1. We give a short justification of why we pick each variable below, as well as a description of how they are calculated from ERA5. There are clearly other variables that we could have considered, but those chosen were selected to cover a broad range of dynamic and thermodynamic environments and are consistent with the choices made in other similar studies (B. Chen et al. 2017, X. Chen et al. 2023; Galarneau et al. 2023).

Any reanalysis product will have biases, due to its underlying model, the observations that have been assimilated, etc. However, the biases in ERA5 are in general reduced compared to those in ERA-Interim (Hersbach et al. 2020), particularly over the analysis period of this study. When compared to observations that have not been assimilated into either product, ERA5 has smaller biases than MERRA-2 for convective environments compared to rawinsondes over Europe and North America (Taszarek et al. 2021) and similarly for tropospheric moisture when compared to satellite-derived observations over the tropics and subtropics (Johnston et al. 2021). However, Virman et al. (2021) did find an increase in the temperature and moisture bias of ERA5 versus ERA-Interim compared to radiosondes over tropical oceans, with ERA5 being too cold and too moist in the low to midtroposphere. While systematic ERA5 biases should be borne in mind, previous work supports ERA5 being a suitable reanalysis product for investigating the environments of MCSs, particularly as we are performing spatial averaging, which will decrease the impact of unbiased noise.

1) TOTAL COLUMN WATER VAPOR

Many studies have shown the importance of TCWV for convection (e.g., Bretherton et al. 2004; Holloway and Neelin 2009, 2010), due to its key role in providing an energy source through the release of latent heat. It is important for the formation of MCSs (Peters et al. 2009) and is a key predictor of MCS precipitation (Chen et al. 2017; Schiro et al. 2020; Chen et al. 2023). It is one of the key ingredients of convection that drives MCSs (Schumacher and Rasmussen 2020). We also consider changes in TCWV over a 3-h period, calculated by subtracting the TCWV field 3 h previously from the current field, as this may indicate that the environment is changing in a way which promotes MCS formation. This was partly motivated by noting that different latitudes all showed an increase in TCWV but had different absolute values, and hence, using a derivative of TCWV might yield more consistent results across different regions.

2) CAPE

Atmospheric instability is one of the key ingredients for MCS (Schumacher and Rasmussen 2020), and CAPE is an often-studied measure of instability. Some studies have shown that changes in instability are a useful precursor condition for MCS (Galarneau et al. 2023). However, other studies found that it was not useful (Chen et al. 2017). Due to its near ubiquitous use in convection parameterization schemes, it is clearly of interest here. We use the standard ERA5 CAPE, calculated as the maximum CAPE of any parcel released from the surface to 350 hPa. It is calculated as the vertical integral of the buoyancy of a pseudoadiabatic parcel ascent without mixing with the environment, from the release layer to the level of neutral buoyancy. As with TCWV, we also consider 3-h changes in CAPE in an attempt to reduce the possible dependence on typical background values at a given latitude.

3) CIN

The presence of atmospheric instability is necessary for deep convection, but if there is too much CIN, the instability

may never be released. CIN is therefore a potential predictive variable for the formation of MCSs (Chen et al. 2017). We note that in Chen et al. (2017), they found that it was not a useful predictor. We use the standard ERA5 CIN, calculated as CAPE but for when the parcel ascent is less buoyant than the environment (here, defined as positive).

4) MOISTURE FLUX CONVERGENCE

Moisture flux convergence (MFC) is defined as

$$\text{MFC} = \int_{p_s}^{p_t} \nabla \cdot (\rho q \mathbf{u}) dp, \tag{1}$$

where the vertical integral is bounded by the surface pressure p_s and the pressure at the model top p_t, ρ is the air density, q is the specific humidity, and $\mathbf{u}$ is the horizontal wind vector. Moisture flux convergence supplies moisture to an MCS or convective region (Schumacher and Rasmussen 2020). The convergence aspect is related to the lift ingredient required for convective activity (Schumacher and Rasmussen 2020). MFC has been used as a closure for convection parameterization schemes (Kuo 1974), and more recently, the related vertically integrated moisture tendency has been used (Becker et al. 2021). We perform a full 3D calculation from the base ERA5 fields. We note that MFC could be low even with large wind convergence in areas of uniformly high moisture. Decomposing MFC into convergence and moisture anomaly could yield insights into which of these are more important in future work, with preliminary results (Figs. S3 and S4) indicating that convergence is more important.

5) SHEAR BETWEEN DIFFERENT HEIGHTS

Vertical wind shear is often a key driver of MCS formation (e.g., Thorpe et al. 1982; Rotunno et al. 1988; LeMone et al. 1998; Schumacher and Rasmussen 2020; Muetzelfeldt et al. 2021a). Low-level shear has been shown to have predictive value for the formation of MCSs (Chen et al. 2017), as has deep shear (Chen et al. 2023). We thus consider multiple measures of shear, defined as the magnitude of the vector wind difference between different levels (these are technically wind differences, not shears, but we stick with the customary usage). The levels are similar to those used in the previous studies and are shown in Table 1. A low- to midlevel shear variable was also considered, but it produces similar results to the other shear variables and is highly correlated with them in the precursor analysis (Fig. S1), so it is not shown.

6) RELATIVE HUMIDITY AT DIFFERENT HEIGHTS

Here, we use the same levels for averaging as defined in Chen et al. (2017), translated into ERA5 model levels. The averaging levels for the two RH variables are shown in Table 1. RH captures information about how much moistening would be necessary to bring the air to saturation and form clouds. The RHlow does this for the boundary layer and thus is indicative of how readily convective clouds can form (Chen et al. 2017). Observational and modeling studies have shown that, over ocean, there is a strong suppression of convection if midlevel

RH is low (Brown and Zhang 1997; Derbyshire et al. 2004), and entrainment of dry air is given as a key mechanism. However, previous work also suggests that MCSs can persist in regions of low–midlevel RH (Zipser 1977; Houze 2004). Therefore, we will investigate RHmid to characterize its effect on MCS activity.

7) EQUIVALENT POTENTIAL TEMPERATURE AT MIDLEVELS

From Chen et al. (2023), their Fig. 1a, a clear maximum in the correlation between θ_e and total MCS lifetime precipitation is seen at the midlevels, with the strongest values between 850 and 750 hPa, indicating the layer over which θ_e is most important. We use the same layer in our analysis of θ_e.

c. Deriving MCS environments

The MCS tracking dataset CCS masks are on a finer grid than ERA5. Therefore, we upscale (coarsen) the MCS cloud masks to the ERA5 grid. As the MCS tracking dataset is stored on the half hour, whereas ERA5 data are stored on the hour, we temporally linearly interpolate the ERA5 data to the half hour.

Figure 1 shows an example of the different regions considered in relation to MCSs, overlaid on the ERA5 TCWV field. We use two methods for determining MCS environments. The first method takes the centroid location of MCSs from the tracking dataset (green dots). From these, a circle of a given radius (green circles) defines a local MCS environment at the given scale, and we take the mean environment defined in these circles. The deep convection initiation (DCI) point is defined as the first cloud which is tracked in the MCS tracking dataset (Feng et al. 2021) and which goes on to form an MCS (i.e., satisfy the area and other criteria). Thus, it captures some of the smaller-scale features that will grow into an MCS. By using the DCI instead of the MCS initiation, we hope to capture more information about the environment before the MCS has started to affect it. (Some literature refers to when an MCS first reaches mesoscale dimension as "genesis," e.g., Coniglio et al. 2010.) Before DCI, we take the initial cloud centroid location and define a precursor environment from circular regions defined by the DCI centroid earlier in time. After DCI, we use the centroid location defined by the cloud track as the MCS forms and propagates. This technique allows us both to define a sensible MCS environment before the MCS has formed and to see how the results depend on the spatial scales of analysis.

The second method uses information on the MCS tracking grid (interpolated to the coarser ERA5 grid), using the defined MCS CCS (shield) and non-MCS CCS. In addition, a convective subregion is defined by a cloud core (as in Feng et al. 2021) of $T_b < 225$ K, for both MCS and non-MCS CCSs. The environmental conditions can then be sampled in five mutually exclusive subregions: MCS core, MCS shield, non-MCS core, non-MCS shield, and a high-cloud-free environment. (A second definition for core subregions was considered, using a precipitation threshold of >2 mm h^{-1}, which has a minor quantitative effect on the results but leads to the same
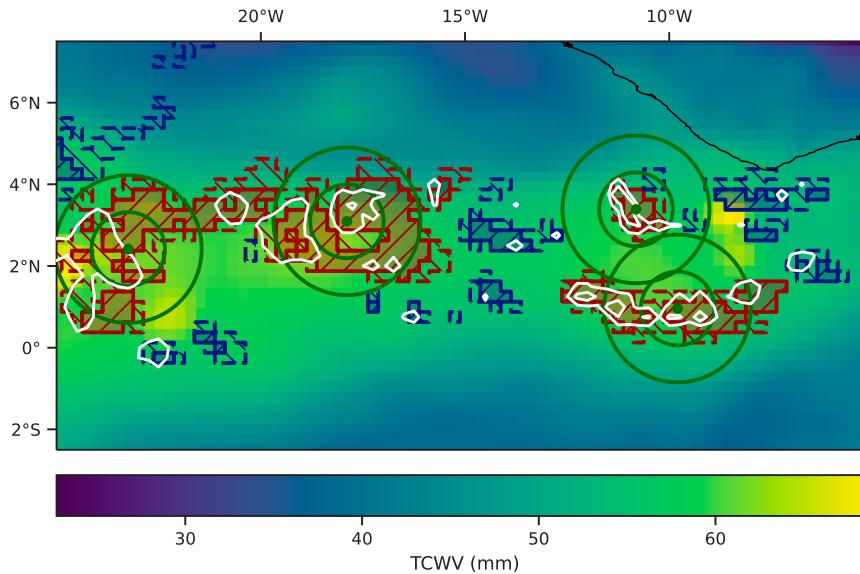
FIG. 1. MCS activity and corresponding environment off the west coast of Africa at 0630 UTC 1 Jan 2020, shown to illustrate the different regions used in subsequent analysis. TCWV (colors) on the ERA5 grid, as well as IMERG precipitation (white contours at 3 and 10 mm h^{-1}). Green dot and circles: MCS centroid and 100- and 200-km radii (500 and 1000 km not shown). Red dashed—MCS cloud shield; red solid—MCS cloud core. Blue dashed—non-MCS cloud shield; blue solid—non-MCS cloud core. In both cases, core and shield are defined by $T_b < 225$ K and $T_b < 241$ K, respectively. Both the core and shield regions have been interpolated to the ERA5 grid.

qualitative conclusions.) The primary motivation for splitting the domain into separate regions is so that we can compare the MCS and non-MCS core subregions. However, to compare against a high-cloud-free environment, we keep the same threshold for the non-MCS shield as it is used for the MCS shield. Thus, the non-MCS shield will include some areas which are not deep convection–related shields (e.g., cirrus clouds associated with midlatitude jets), but we stick with this naming. Both non-MCS subregions have no area threshold—they can include contributions from a single pixel. This method allows us to sample the MCS (and non-MCS) environmental conditions exactly where the MCS is found and furthermore defines a core convective subregion for further analysis. However, it is not possible to use these definitions to investigate the precursor environment nor does it allow easy analysis of spatial scales. Thus, it is complementary to the previous method.

3. Composite MCS precursor and contemporaneous environments

To investigate the changes in the environment before and during MCSs, we examine the environments defined by circular regions centered on the MCS as shown in Fig. 1. We only consider MCSs that have not formed as the result of a split from an existing MCS, as this ensures that the precursor environment does not contain an MCS. This decision results in strengthening of the signal seen in the evolution of some variables before DCI (see Fig. S2), although some individual

events might be lost (e.g., a supercell forming from an existing MCS). This results in 366 101 MCSs being included in our analysis—having a large sample size means we can decompose our analysis in various ways and still have reasonable representation in each grouping. We show the composite environments, defined as the mean over the circular regions at a given time, with the average then taken over all MCSs in a given geographical domain.

Figure 2 shows the precursor and contemporaneous values for 12 selected environmental variables, separated by geographical domains based on latitudinal bands and whether the MCS occurs mainly over land or sea, with the analysis carried out over a scale of 200 km. The land–sea split is defined by categorizing any MCS that has over a 50% mean land fraction as occurring over land. The latitude is defined by the mean latitude of the MCSs. The equatorial latitude band has the most MCSs, with 47.7% of MCSs occurring there. The tropics have a larger fraction of MCSs than the extratropics, and all latitude bands have higher occurrence over sea than over land with approximately 50% more. This is consistent with global MCS distributions (Laing and Fritsch 1997; Feng et al. 2021). Most variables have some latitudinal dependence and marked differences between land and sea. In general, there is a small difference between the equatorial latitude band and the tropics and a larger difference between both of these and the extratropics.

Focusing on CAPE (Fig. 2a), there is a clear difference in behavior between land and sea, with land-based MCSs showing a large increase in CAPE before initiation over all latitudes.
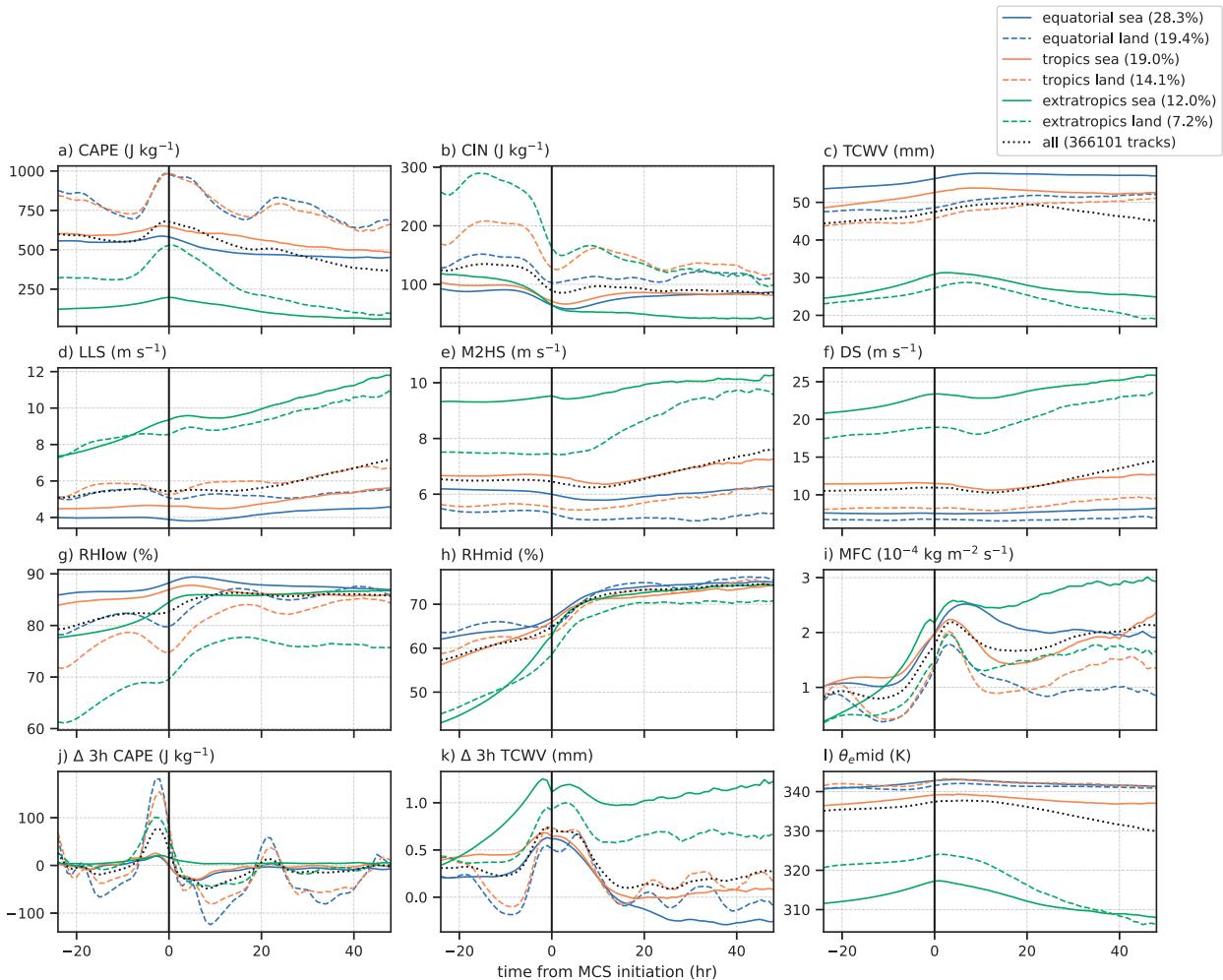
FIG. 2. Composite environmental conditions relative to DCI time, separated by latitude bands (equatorial: 10°S–10°N, tropical: 10°–30°, extratropical: 30°–60°) and land–sea, as well as all tracked MCSs. Before initiation (black vertical line), the precursor environment is calculated by taking the mean environment at a radius of 200 km centered on the location of the DCI centroid (see Fig. 1). After initiation, the radius is centered on the centroid of the MCS as it moves. The proportion of tracks in each latitude band is indicated in the legend.

The absolute value of CAPE is a function of latitude, with the highest peak values over land of around 1000 J kg^{-1} in the equatorial and tropical latitude bands. There is a peak in CAPE around the initiation time. Over equatorial and tropical land, there are minima around 16 h after initiation, which is due to the diurnal cycle of CAPE. From the change in CAPE over 3 h (Fig. 2j), the land–sea difference is again pronounced, with the largest rate of change seen for the equatorial latitude band over land. This variable exhibits reduced latitudinal dependence, which was a motivation for calculating the change in CAPE [section 2b(2)]. The change in CAPE is far smaller over sea than over land, but there is nonetheless a weak increase over sea before initiation. CIN (Fig. 2b) shows a marked decrease before initiation. CIN increases again after initiation, although not to the same levels as seen before initiation.

TCWV (Fig. 2c) shows a strong latitudinal dependence, with higher values in the lower latitudes. There is also a clear difference between TCWV over sea and land at lower

latitudes, with the sea regions being moister. There is an increase in TCWV before initiation for all regions, although it is not as pronounced as that of CAPE. However, when the 3-h change in TCWV is considered, a clear and consistent signal is found, starting around 6 h before initiation (Fig. 2k). The closely related MFC (Fig. 2i) is remarkable in that it has a similar precursor behavior over all regions, with a marked increase before initiation reaching a mean value between 1.4 and 2.1 × 10^{-4} kg m^{-2} s^{-1}. For this reason, we decompose MFC by both diurnal cycle (Fig. 4) and season (Fig. 5) below. There is a strong variation in MFC after initiation, with land MCSs showing a sharp decrease after a peak a couple of hours after initiation, whereas sea MCSs show a less pronounced peak.

The shear variables show relatively weak increases, or even decreases, before initiation (Figs. 2d–f). Low-level shear (LLS) in particular shows weak dependence on land–sea. Over the extratropics, LLS shows the clearest increase before
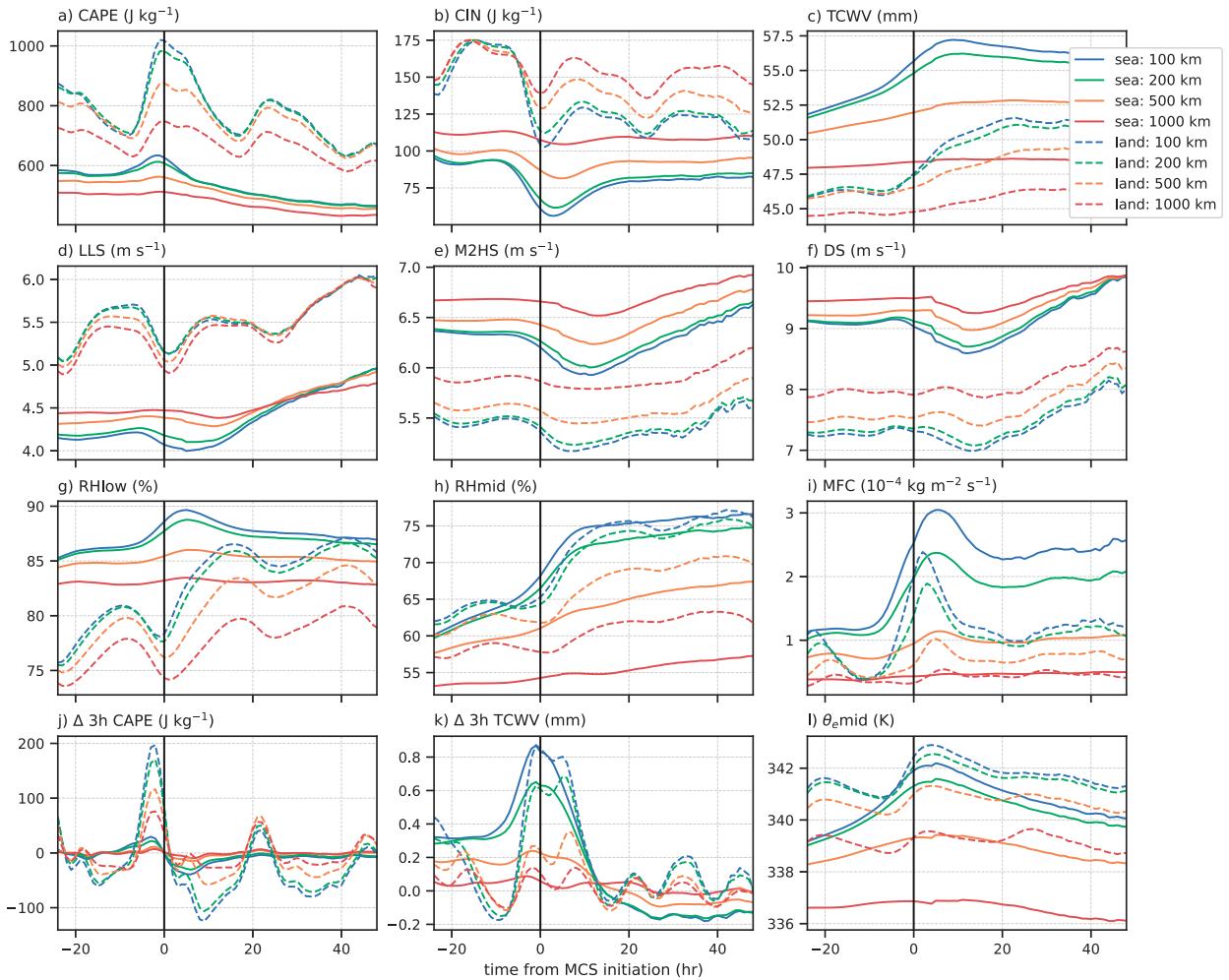
FIG. 3. As in Fig. 2, but with the analysis carried out at four radii and for equatorial and tropical MCSs only (30°S–30°N) and again separated by land–sea.

initiation. Over all latitude bands, LLS exhibits interesting behavior where it increases after initiation (weakly so over equatorial sea). This could be due to two things. First, the MCSs must last for at least 4 h, but fewer and fewer will last beyond this (see Fig. S5). Thus, it is possible that in these regions, MCSs which last for longer have higher sustained values of LLS. Second, it is possible that there is a countergradient momentum transport (e.g., LeMone et al. 1984) associated with the MCS, which is acting to strengthen the LLS. Similar remarks on shear strengthening apply to M2HS and DS.

RHlow (Fig. 2g) shows a strong dependence on land–sea, with land MCSs having a distinct structure of RHlow environment. Over land, all latitude bands show a peak at around 10 h before initiation, a minimum at initiation (weakest over the extratropics), with a secondary peak around 16 h after initiation. Over sea, there is far less variation due to the lower levels being close to saturation. RHmid (Fig. 2h) shows very little land–sea dependence, indicating the midtroposphere is much less affected by the surface. There is a consistent signal of an increase in RHmid at all latitudes, particularly after

initiation, probably due to small-scale convection in the precursor environment acting to moisten the midtroposphere and making the conditions more favorable for larger-scale convection to occur. The $\theta_e$mid (Fig. 2l) does not display much change, except over higher latitudes, where there is a weak increase before initiation and a weak peak at or shortly after initiation.

a. Spatial scales

Investigating the precursor environments over different spatial scales can reveal insights about which scales are important for each variable. Because most of the MCSs occur over the tropics and equatorial latitude bands (81%), and these regions are similar in terms of their environments (Fig. 2), we limit our analysis to these latitude bands combined. Figure 3 shows the 12 environmental variables, split over land–sea, capturing the environment at four spatial scales: 100, 200, 500, and 1000 km. Different variables have a different dependence on the spatial scale of analysis. For example, LLS shows little dependence on scale over land or sea (Fig. 3d). However, TCWV displays
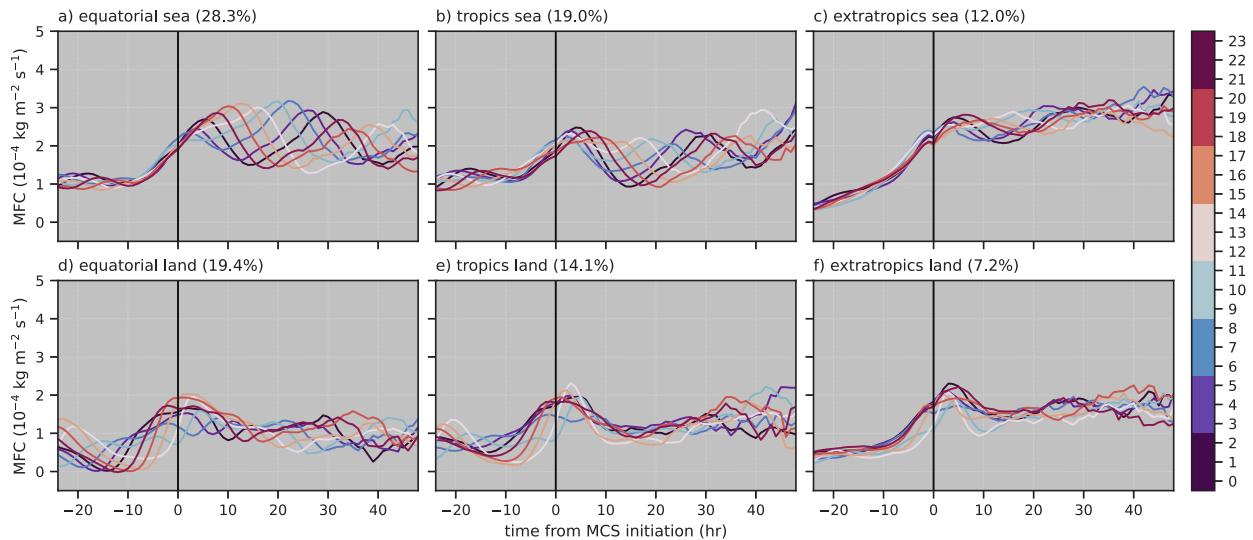
FIG. 4. MFC shown for each geographical region (as in Fig. 2) and decomposed by the diurnal cycle (as shown by the color of the line, with the color bar providing a key). The minimum number of MCSs in any individual category is 1791 for extratropical land in the bin at 0300–0500 LST, indicating that each category has a reasonable number of samples in it.

qualitatively different behavior at different spatial scales, with a clear change in gradient before initiation seen at 100 and 200 km over sea and no change visible at 1000 km (Fig. 3c). In general, the relative changes are greatest at the smallest scales, and the changes close to initiation are sometimes lost when averaging over a larger area.

Some variables show a different dependence on scale over land than over sea. For example, at 1000 km, CAPE and CIN show a smaller change over sea (Figs. 3a,b), but a marked change over land, indicating that the largest spatial scale analyzed here is important over land but not over sea. RHmid and $\theta_e$mid also vary over larger spatial scales over land than over sea (Figs. 3h,l). Similarly, LLS and RHlow (Figs. 3d,g) show a large change over land over all spatial scales.

For each of the variables and over land and sea, the behavior at 100 and 200 km is similar (i.e., the blue and orange lines, solid and dashed, are close to each other for each variable), whereas there is often a marked change between these and the 500-km scale (orange lines). This potentially indicates that there is stronger scale dependence in the environmental forcing between 200 and 500 km. It could be that the gaps between 100 and 200 km and 100 and 500 km are explained by a linear relationship between the analysis scale and the size of the gap. This was tested (not shown), and this is not the case for certain variables, notably TCWV, RHmid, and $\theta_e$mid, indicating that the scale dependence is indeed stronger above 200 km for these.

b. Diurnal and seasonal cycle of moisture flux convergence

From Fig. 2i, MFC presents a consistent evolution in the precursor environment. We further decompose MFC by diurnal cycle in Fig. 4, where the initiation time in local solar time (LST) is used for eight 3-h groups. For the diurnal cycle of MFC over sea (Figs. 4a–c), there is very little variation in the

precursor environment. After initiation, particularly over the tropics and extratropics, the diurnal cycle becomes more important, with morning MCSs (0600–0800 LST) peaking soonest at around 3 h after initiation, and early afternoon MCSs (1200–1400 LST) peaking latest at around 18 h after initiation. Indeed, for each diurnal cycle group, there is a peak at around 1000 LST, which is consistent with MCS activity over sea peaking in the morning (e.g., Mohr and Zipser 1996; Feng et al. 2023b). There is a stronger diurnal cycle at precursor times over equatorial land (Fig. 4d), but it is still very weak compared to most other environmental variables (cf. with CAPE and CIN, which have a strong diurnal cycle—see Figs. S6 and S7). The range of values over all regions and for all initiation LSTs is 0.9–2.3 × 10^{-4} kg m^{-2} s^{-1}, with extratropical sea MCSs occurring at the top of this range. Considering the uncertainty by including the 25th and 75th percentiles (Fig. S8), the signal is clearly represented in the percentile bounds as well, and there is little difference between the different diurnal cycle groups compared to the variability within each group. In summary, the evolution of the precursor environment of MFC is particularly consistent as a function of both location and diurnal cycle.

We further show the seasonal dependence of MFC in Fig. 5. This shows far less variation than is caused by the diurnal cycle over all regions and seasons. It may be tempting to use the results concerning MFC above to argue that an increase in MFC leads to MCS initiation. However, this is not necessarily the case. The results shown so far indicate that when an MCS forms, the environment can be expected to change in a certain way before its formation. It is perfectly possible that the environment could show the same change before the outbreak of scattered deep convection only, which means that using this change to predict MCS activity could lead to false positives.
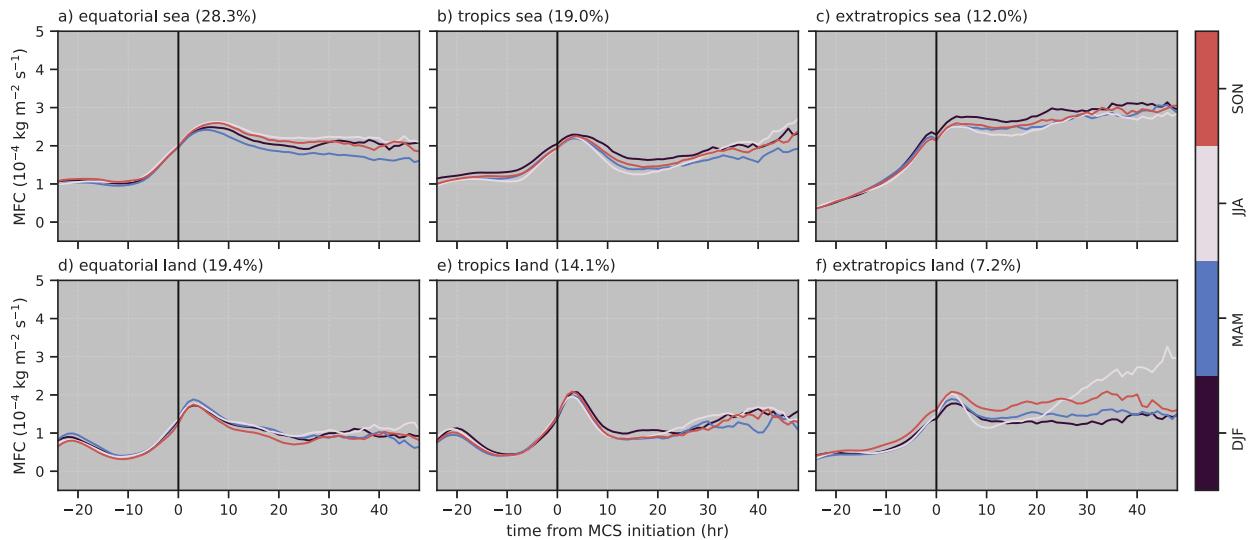
FIG. 5. As in Fig. 4, but decomposed by season.

c. Regional variation in MCS initiation environments

When studying MCSs over a particular geographical domain, a common technique is to produce a composite field for when an MCS initiates (e.g., Li et al. 2023b). Here, we generalize the technique, by averaging fields within 500 km of an initiation event. To take into account the variations in each field over the course of a year, we take the anomaly of the spatially averaged value for each initiation event with respect to the local monthly mean of that field. In Fig. 6, we present the composite anomalies for the 12 analyzed fields. The 500-km averaging scale balances the need to capture information from near the initiation point with the necessity to produce a valid mean over large areas. The other averaging scales show broadly similar features, although as the scale increases, the magnitude of the anomaly decreases (not shown).

Figure 6 shows that the composite anomalies are consistent with the results from Fig. 2 (i.e., $t = 0$) over most of the globe. For example, the CAPE anomaly is positive over 91% of grid points, and the CIN anomaly is negative over 86% of grid points, consistent with Figs. 2a and 2b. CAPE and CIN both
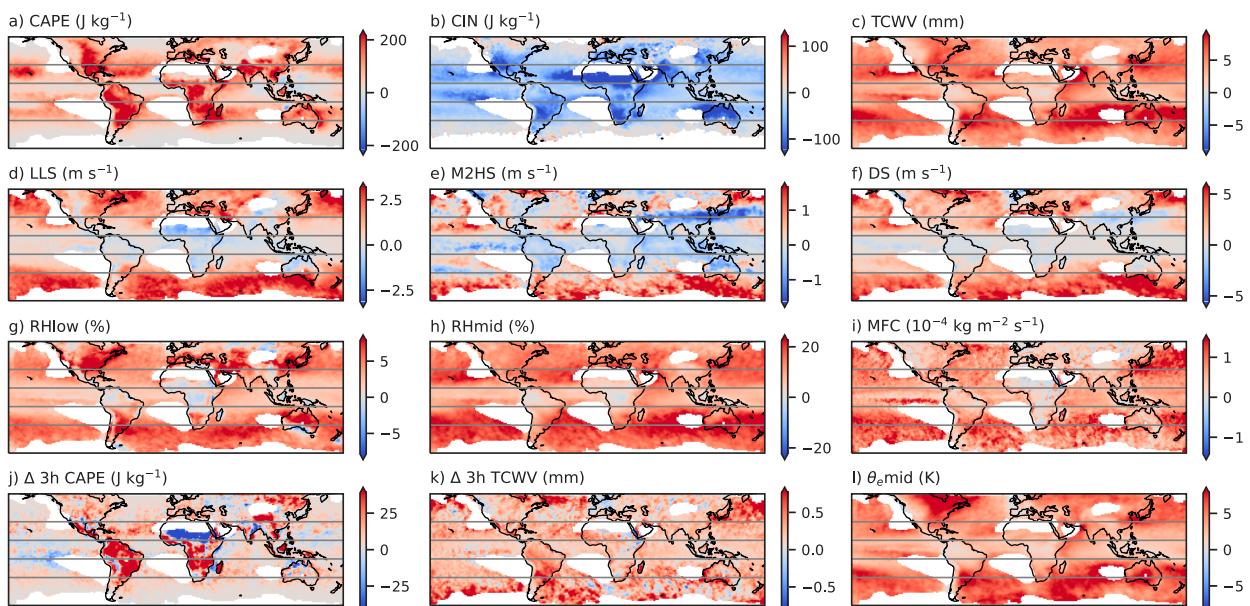

FIG. 6. Composite environmental condition anomalies at MCS initiation calculated over a 500-km length scale, as described in the main text. Grid points are masked out (white) if they have fewer than 100 initiation values. Latitude markers are shown at 10°N, 30°N, 30°S, and 10°S, shown to enable easier comparison with Fig. 2 above.

have a larger anomaly over land, again consistent with Fig. 2. TCWV, RHmid, and $\theta_e$mid all have positive anomalies over more than 99% of grid points, with CAPE, RHlow, and MFC all having positive anomalies over greater than 90%, indicating generally robust anomalies over different regions. Only M2HS and the Δ 3-h CAPE have less than 80% agreement in the sign of the anomaly. Some of the noise at higher latitude can be explained by the fact that there are far fewer MCS counts at these locations (Fig. S9).

There are some noteworthy regional differences in the sign of the anomalies. M2HS (Fig. 6e) shows mainly positive values over higher latitudes and negative values at lower latitudes. This is not evident from Fig. 2e, where there is little increase at $t = 0$, suggesting that the removal of the seasonal cycle may be important for this variable, although inspection of the seasonal cycle (not shown) for this variable was not conclusive. The negative M2HS anomaly is particularly strong over eastern China and northern India, as well as being negative over large regions of the tropics where MCSs often occur, indicating that lower-than-normal M2HS is associated with MCS formation. TCWV has a strong anomaly over the subtropical Southern Hemisphere sea, which is consistent with Fig. 2c, although the strength of the signal in this region is stronger than might be expected from Fig. 2c alone. This could be due to the subtropics from both hemispheres being treated together or due to the removal of the monthly mean in calculating the anomaly.

d. Discussion of MCS composite environments

Many variables show a marked evolution before initiation, with the strongest changes found over land. This is similar to the results of Galarneau et al. (2023), although they perform their analysis by computing changes relative to a randomly sampled reference background, and their study only includes MCSs from one tropical season, September–November. Specifically, they find that a measure of instability (defined as the layer-mean θ_e of 1000–700 hPa minus 700–300 hPa) shows an increase over land around 6 h before MCS initiation (their Fig. 3b). This is closely related to our finding that CAPE shows a marked increase over land, and the timings of the increase are consistent. Galarneau et al. (2023) also find an increase in differential divergence, defined as the divergence at 200 hPa minus the divergence at 925 hPa. This is most closely related to our MFC variable and is also consistent with our findings on MFC, although they do not perform diurnal or seasonal cycle decomposition.

The consistency of MFC changes before initiation stands out. However, there is clearly a question of causality at play (Mapes 1997; Yano and Plant 2012). Does the presence of deep convection before the DCI lead to ascent (such as scattered deep convection or convection over too small area to be detected in the satellite brightness temperature), which in turn leads to a positive MFC? Or does the positive MFC, and its sharp increase before DCI, lead to the development of convection followed by the formation of an MCS? It is also possible that both processes mutually reinforce each other,

leading to a positive feedback and upscale growth resulting in an MCS.

A natural question to consider is whether the changes in MFC are driven by changes in moisture, changes in convergence, or both. In Fig. S3, we show a similar figure to Fig. 3, but for 2020 only, and including MFC, RHlow, RHmid, and six measures of convergence. These are the horizontal convergence at 850, 700, and 600 hPa, as well as the vertical integrals of the convergence from the surface up to each of these levels. It is clear that there is a high degree of similarity between the MFC results and those for convergence at 850 hPa and even higher similarity for the integral of convergence up to 700 hPa, whereas there is little similarity between the MFC results and those for either RH variable. This picture is confirmed by the correlations between MFC and the aforementioned variables (Fig. S4), which indicate that there is little correlation between MFC and the RHs at DCI but relatively high correlation between MFC and the integral of convergence up to 700 hPa ($r^2 = 0.45$).

The results of this section are of interest in their own right. However, they may also prove useful for parameterizing MCS effects in global models. First, the evidence of scale dependence in environmental indicators is relevant for consideration of the length scales over which a parameterization should be active. Second, if it is possible to track MCSs in the global models, then the same analysis as performed here could be applied to determine whether the environment changes in the same manner before and after simulated MCSs. Discrepancies would point toward deficiencies in the models' representation of MCSs and the environments under which they form. For example, MCSs have recently been studied in the Dynamics of the Atmospheric General Circulation Modeled on Non-Hydrostatic Domains (DYAMOND) high-resolution global simulations (Feng et al. 2023b) and a lower-resolution global model (Dong et al. 2021). It would be possible to analyze the environmental conditions of these simulated MCSs, to establish whether differences between observed and simulated MCSs are attributable to differences in the representation of the environment in which they form.

4. MCS occurrence conditional on environment

We now turn to the question of how to predict the likelihood of MCS occurrence for a given value of an environmental variable. We first construct probability density functions (PDFs) of each variable over the entire domain and compare them to the corresponding distributions over cloudy subregions. Specifically, we consider the MCS subregions illustrated in Fig. 1: MCS core, MCS shield, non-MCS core, non-MCS shield, and high-cloud-free environment. We refer to all core and shield subregions as the "cloudy" subregions below. The five subregions are mutually exclusive (e.g., MCS shield is not a superset of MCS core but rather a separate subregion). Together, they completely cover the domain. All PDFs are based on the entire lifetime of MCSs.

PDFs of TCWV, RHmid, and MFC in each of the subregions are shown in Fig. 7 top row. These three variables have been chosen because they display the most contrasting
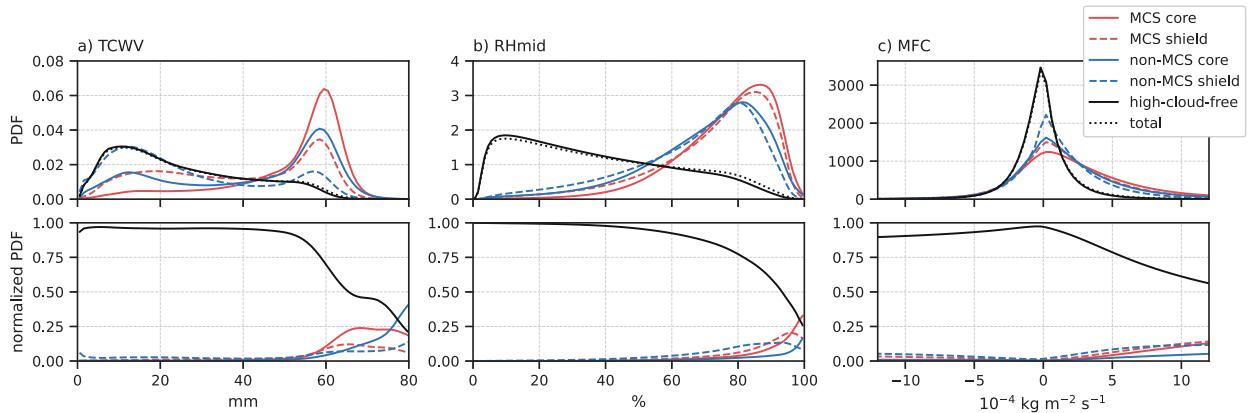
F_{IG}. 7. (top) Global PDFs of environmental variables within different subregions and for all subregions combined ("total"). (bottom) The PDFs normalized taking into account the area of each subregion (note that the high-cloud-free subregion has a far larger area than the others), giving the probability of being in a particular subregion for a given value of environmental variable. Subregions are as in Fig. 1.

behaviors among different subregions (see also Fig. S10 for other environmental variables). The distributions of each variable are easily distinguishable between the different subregions. TCWV shows hints of a bimodal structure, consistent with previous studies (Zhang et al. 2003; Mapes et al. 2018), with modes at around 10 mm for the high-cloud-free subregion and 60 mm for the cloudy subregions. However, the total PDF does not show two distinct modes because we have applied the analysis globally and not just over the tropics, as in previous studies. Each cloudy PDF is different from the others, with MCS core having the highest modal value of TCWV, followed by non-MCS core, MCS shield, and non-MCS shield.

The PDFs for all cloudy subregions for RHmid are heavily skewed toward 100%, although the differences between core and shield are less pronounced than for TCWV. For MFC, the cloudy PDFs are skewed toward the right, indicating that all cloudy subregions are more likely to have positive values of moisture convergence.

As illustrated in Fig. 1, the high-cloud-free subregion covers a much larger area than any of the cloudy subregions. To answer the question of how likely it is to be in a particular subregion for a given environmental value, it is necessary to take into account the area of each subregion. This probability is shown in the bottom row of Fig. 7 (and Fig. S10). For TCWV, there is over a 90% chance of being in the high-cloud-free subregion until a value of 53 mm is reached, and above which the probability drops sharply to 50% by 65 mm. Above 58 mm, the next highest probability is that of the MCS core subregion, which increases more quickly than any other cloudy subregion. The sharp drop above 50 mm is related to the idea of a critical value of water content, as described by Peters et al. (2009) for example.

Large values of RHmid and MFC also show decreases in the probability of being in the high-cloud-free subregion, dropping to 30% for RHmid when the midlevels are completely saturated and to 60% for MFC for very large moisture convergence. However, they do not display an analogous critical value

behavior, in that they do not show a marked change in gradient at a particular value. CAPE and CIN display little change of probability as a function of the variable value (Figs. S11a,b) and so are not shown here. Indeed, they are poor predictors of subregion for this reason. The three variables shown in Fig. 7 have the lowest probabilities of being in the high-cloud-free subregion for variable values that are commonly visited by the full PDF and therefore the highest relative predictive power of all variables, particularly of the probability of not being in the high-cloud-free subregion.

MCS convection conditional on any convection

The results of the previous subsection provide a predictive model for being in a particular MCS (or non-MCS) subregion for a given value of environmental variable. Here, we further build on that idea by refining the question to: Given that convection is active (as determined by being in either an MCS or non-MCS core subregion), what is the probability of it being MCS-type convection? Thus, we calculate p(MCS convection|convection), as a function of the environmental variable. Figure 8 shows this probability for the three chosen variables, as well as the full PDF for each variable. The probabilities for the entire lifetime of all MCSs are calculated over the entire domain and over land and sea individually. There are some differences between land and sea, but broadly speaking the distributions have similar structures for all three variables. TCWV and RHmid show a clear increase in the probability of being in an MCS convective subregion given that convection is active as the variable value increases. This marks them out as being good candidates for providing a predictive relationship. Indeed, in both cases, a simple linear relationship would be practical for estimating the probability. The conditional probability, p(MCS convection|convection), for MFC does not show nearly as much variation over land, sea, or all regions (particularly where the PDF is non-negligible), indicating it is not as useful for predicting this probability.

Figure 9 refines the above analysis into three latitudinal bands: equatorial (10°S–10°N), tropics (10°–30°), and extratropics
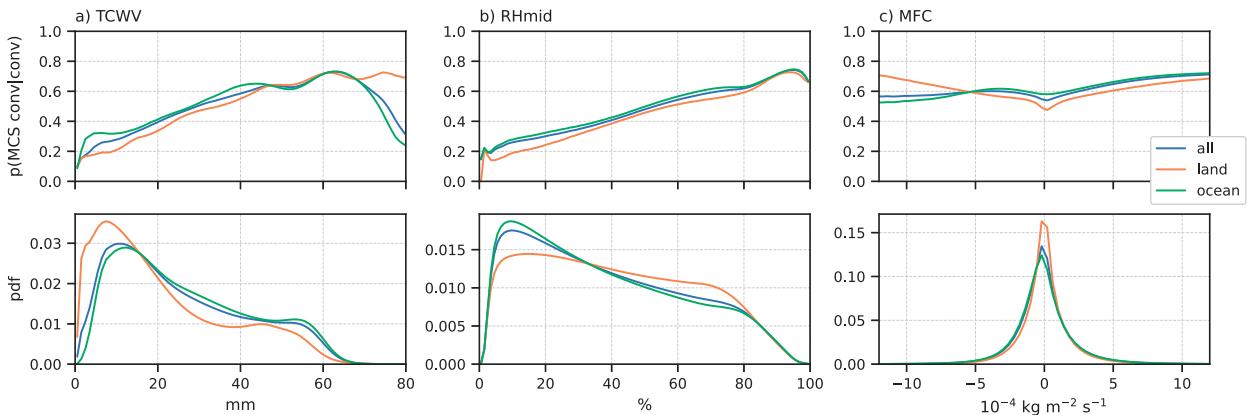
FIG. 8. Conditional probability of p(MCS convection|convection), calculated by dividing the MCS convection in Fig. 7 (bottom) by the sum of MCS and non-MCS convection. (top) Conditional probabilities over the entire domain, over land, and over sea. (bottom) Complete distribution of each variable $p(x)$ (i.e., taking into account all five subregions) to indicate coverage of each variable's distribution.

(30°–60°). The distribution for RHmid is very consistent across the low-latitude bands (Fig. 9b) but shows a weaker relationship in the extratropics. TCWV (Fig. 9a) shows an increasing likelihood of being in an MCS-type convective subregion for all latitude bands above around 30 mm. However, the absolute likelihood is highest for the extratropics, and below 30 mm, there is an increase in the likelihood of being in an MCS-type convective subregion in the equatorial latitude band. MFC (Fig. 9c) shows the same shape of distribution over equatorial and extratropical latitudes, although at a higher probability throughout over the extratropics. There is a weaker response over the tropics.

5. Discussion of conditional MCS occurrence

The results here provide the basis for a predictive model of MCS occurrence given a background environmental state. Major features of the results are robust over land–sea and over different latitudes. Our results are also not sensitive to diurnal or seasonal cycle (Fig. S12). They have similarities to those in Chen et al. (2017), in that TCWV and RHmid emerge as important variables, although we find that the shear variables do not have the same importance. However, Chen et al. (2017) were specifically concerned with MCS rainfall, whereas we are focused on MCS occurrence, as this fits our follow-on application.

Unlike Chen et al. (2017), we distinguish MCS from non-MCS convection, which allows us to discern the differences between these modes of convection and the environments

that give rise to them. Furthermore, our results are not overly sensitive to the definition of a convective core: a definition based on precipitation, using a threshold of 2 mm h^{-1} yields similar results (see Fig. S13).

This analysis has value in its own right, but its full importance becomes clear when considering the parameterization of MCSs. In a global climate model, convection is parameterized. If we assume that the convection scheme correctly identifies active convection in its model, then this can be matched to the observational condition above that convection is active. Our results then allow us to assign a probability that the parameterized convection should have MCS or non-MCS characteristics based on the environmental conditions. Thus, the conditional probability could be used as a probabilistic trigger condition for use by an MCS parameterization scheme.

6. Summary

By combining a near-global long-term MCS tracking dataset with ERA5, we have performed a statistical investigation of the environmental conditions under which MCSs form and persist. Certain variables are found to change before MCS initiation when defined as averages within circular regions centered on the MCS location. For example, CAPE increases before initiation and CIN decreases, with the signal being strongest over land. MFC shows a remarkably consistent increase before initiation across different latitude bands and over both land and sea, as well as when decomposed by diurnal or seasonal cycle. By performing the analysis at different
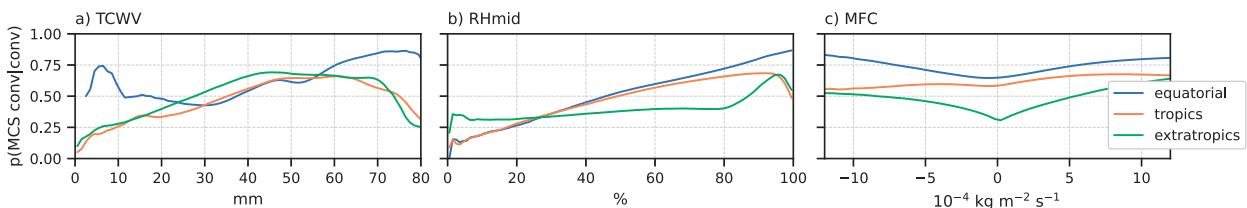

FIG. 9. As in the top row of Fig. 8, but separated into three latitudinal bands (defined in the main text).

spatial scales, a change in behavior for some variables between 200 and 500 km is identified (section 3a). The smaller of these scales, 200 km, corresponds to the scale over which the environment is important for the formation of MCSs. Such information could prove useful for the parameterization of MCSs, setting a natural scale over which it is important to consider the environment for the triggering of an MCS scheme.

A complementary statistical analysis is performed by splitting the domain into five mutually exclusive subregions: MCS core, MCS shield, non-MCS core, non-MCS shield and high-cloud-free environment, and determining the probability of occurrence of each region conditional on the environment. Performing the analysis in this way is necessary for both building a statistical model of when MCSs might occur based on environmental conditions and for determining when an MCS scheme should be active in a global model. PDFs of all variables are derived in these subregions. When the relative areas of the subregions are taken into account, these can be used to build a probabilistic model of the likelihood of occurrence of a particular subregion given the environmental state. Three variables stand out as indicating the highest likelihood of being in a cloudy region: TCWV, RHmid, and MFC. Of these, TCWV and RHmid both show promise in being able to discriminate between MCS-type convection and non-MCS-type convection, given that convection is active. This probabilistic information is ideal for devising the probabilistic trigger conditions for an MCS scheme.

We have developed a diagnostic analysis that allows the probabilistic prediction of an MCS to be made given the environmental state. This could be applied to ERA5 environments to back out a prediction of MCS occurrence, on the grounds that ERA5 does not have the resolution to faithfully reproduce an MCS, but it may have sufficient resolution to represent its environment. Furthermore, it could be applied to global models as a proxy for investigating the occurrence of MCSs under climate change, for the same reason as for ERA5.

We have attempted to choose a broad selection of environmental variables which are important for MCS formation and maintenance and have found that TCWV is important. Previous studies (e.g., Bretherton et al. 2004) have shown that a saturated version of TCWV can be useful, as it allows relationships to be uncovered that hold over more global regions. Additionally, entraining CAPE (Peters et al. 2023), a version of CAPE which takes into account entrainment of environmental air by diagnosing a cloud radius from the wind profile, has been shown to perform better at predicting updraft velocities than standard CAPE. Both of these could be included as additional variables in the future. Further variables could also be considered, or certain variables that are important here could be split into constituent parts. For example, MFC could be decomposed into convergence and moisture anomaly terms, with our preliminary work suggesting that the dynamical convergence would be the predominant factor in driving changes in MFC.

We intend to use the information generated here to implement modifications to the MCS parameterization scheme of Zhang et al. (2024, manuscript submitted to *J. Adv. Model. Earth Syst.*), which builds on the earlier work by Moncrieff et al. (2017) and Chen et al. (2021). The characteristic environmental scale could be used to determine on what scales the scheme should be active. The relationships between MCS occurrence probabilities given convection is active as a function of the environmental state can be used as a probabilistic trigger. The three variables that show the clearest relationship here, TCWV, RHmid, and MFC, could be used as inputs to this scheme. Due to TCWV's previous links to deep convection and MCSs (e.g., Peters et al. 2009; Chen et al. 2017; Schiro et al. 2020), and its clear predictive value for MCSs, it is a good candidate to use. Questions would still need to be answered about how well TCWV is represented in global models. For example, Mapes et al. (2018) show that its tropical bimodal structure is not well represented in models. It might be that a combination of environmental variables is preferable to using a single variable, perhaps using multiple linear regression or machine learning techniques, or it might be that the variables are sufficiently correlated that this does not yield further useful information. This should be investigated.

REFERENCES

Ahmed, F., and C. Schumacher, 2015: Convective and stratiform components of the precipitation-moisture relationship. *Geophys. Res. Lett.*, **42**, 10 453–10 462, https://doi.org/10.1002/2015GL066957.

Becker, T., P. Bechtold, and I. Sandu, 2021: Characteristics of convective precipitation over tropical Africa in storm-resolving global simulations. *Quart. J. Roy. Meteor. Soc.*, **147**, 4388–4407, https://doi.org/10.1002/qj.4185.

Bowman, K., and C. Homeyer, 2017: GridRad-three-dimensional gridded NEXRAD WSR-88D radar data. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, accessed 30 January 2025, https://doi.org/10.5065/D6NK3CR7.

Bretherton, C. S., M. E. Peters, and L. E. Back, 2004: Relationships between water vapor path and precipitation over the tropical oceans. *J. Climate*, **17**, 1517–1528, https://doi.org/10.1175/1520-0442(2004)017<1517:RBWVPA>2.0.CO;2.

Brown, R. G., and C. Zhang, 1997: Variability of midtropospheric moisture and its effect on cloud-top height distribution during TOGA COARE. *J. Atmos. Sci.*, **54**, 2760–2774, https://doi.org/10.1175/1520-0469(1997)054<2760:VOMMAI>2.0.CO;2.

Chen, B., C. Liu, and B. E. Mapes, 2017: Relationships between large precipitating systems and atmospheric factors at a grid scale. *J. Atmos. Sci.*, **74**, 531–552, https://doi.org/10.1175/JAS-D-16-0049.1.

Chen, C.-C., J. H. Richter, C. Liu, M. W. Moncrieff, Q. Tang, W. Lin, S. Xie, and P. J. Rasch, 2021: Effects of organized convection parameterization on the MJO and precipitation in E3SMv1. Part I: Mesoscale heating. *J. Adv. Model. Earth Syst.*, **13**, e2020MS002401, https://doi.org/10.1029/2020MS002401.

Chen, D., J. Guo, D. Yao, Z. Feng, and Y. Lin, 2020: Elucidating the life cycle of warm-season mesoscale convective systems in eastern China from the Himawari-8 geostationary satellite. *Remote Sens.*, **12**, 2307, https://doi.org/10.3390/rs12142307.

Chen, X., L. R. Leung, Z. Feng, and Q. Yang, 2022: Precipitation-moisture coupling over tropical oceans: Sequential roles of shallow, deep, and mesoscale convective systems. *Geophys. Res. Lett.*, **49**, e2022GL097836, https://doi.org/10.1029/2022GL097836.

——, ——, ——, and ——, 2023: Environmental controls on MCS lifetime rainfall over tropical oceans. *Geophys. Res. Lett.*, **50**, e2023GL103267, https://doi.org/10.1029/2023GL103267.

Coniglio, M. C., J. Y. Hwang, and D. J. Stensrud, 2010: Environmental factors in the upscale growth and longevity of MCSs derived from Rapid Update Cycle analyses. *Mon. Wea. Rev.*, **138**, 3514–3539, https://doi.org/10.1175/2010MWR3233.1.

Daleu, C. L., R. S. Plant, A. J. Stirling, and M. Whitall, 2023: Evaluating the CoMorph-A parametrization using idealized simulations of the two-way coupling between convection and large-scale dynamics. *Quart. J. Roy. Meteor. Soc.*, **149**, 3087–3109, https://doi.org/10.1002/qj.4547.

Derbyshire, S. H., I. Beau, P. Bechtold, J.-Y. Grandpeix, J.-M. Piriou, J.-L. Redelsperger, and P. M. M. Soares, 2004: Sensitivity of moist convection to environmental humidity. *Quart. J. Roy. Meteor. Soc.*, **130**, 3055–3079, https://doi.org/10.1256/qj.03.130.

Dong, W., M. Zhao, Y. Ming, and V. Ramaswamy, 2021: Representation of tropical mesoscale convective systems in a general circulation model: Climatology and response to global warming. *J. Climate*, **34**, 5657–5671, https://doi.org/10.1175/JCLI-D-20-0535.1.

Enyew, B. D., and A. Mekonnen, 2022: The interaction between African easterly waves and different types of deep convection and its influence on Atlantic tropical cyclones. *Atmosphere*, **13**, 5, https://doi.org/10.3390/atmos13010005.

Feng, Z., 2023: PyFLEXTRKR global MCS tracking dataset using GPM MergedIR Tb and IMERG precipitation data. Zenodo, accessed 30 January 2025, https://doi.org/10.5281/zenodo.10023498.

——, X. Dong, B. Xi, C. Schumacher, P. Minnis, and M. Khaiyer, 2011: Top-of-atmosphere radiation budget of convective core/stratiform rain and anvil clouds from deep convective systems. *J. Geophys. Res.*, **116**, D23202, https://doi.org/10.1029/2011JD016451.

——, and Coauthors, 2021: A global high-resolution mesoscale convective system database using satellite-derived cloud tops, surface precipitation, and tracking. *J. Geophys. Res. Atmos.*, **126**, e2020JD034202, https://doi.org/10.1029/2020JD034202.

——, J. Hardin, H. C. Barnes, J. Li, L. R. Leung, A. Varble, and Z. Zhang, 2023a: PyFLEXTRKR: A flexible feature tracking Python software for convective cloud analysis. *Geosci. Model Dev.*, **16**, 2753–2776, https://doi.org/10.5194/gmd-16-2753-2023.

——, L. R. Leung, J. Hardin, C. R. Terai, F. Song, and P. Caldwell, 2023b: Mesoscale convective systems in DYAMOND global convection-permitting simulations. *Geophys. Res. Lett.*, **50**, e2022GL102603, https://doi.org/10.1029/2022GL102603.

Fritsch, J. M., and G. S. Forbes, 2001: Mesoscale convective systems. *Severe Convective Storms*, *Meteor. Monogr.*, No. 28, Amer. Meteor. Soc., 323–357, https://doi.org/10.1175/0065-9401-28.50.323.

——, R. J. Kane, and C. R. Chelius, 1986: The contribution of mesoscale convective weather systems to the warm-season precipitation in the United States. *J. Climate Appl. Meteor.*, **25**, 1333–1345, https://doi.org/10.1175/1520-0450(1986)025<1333:TCOMCW>2.0.CO;2.

Galarneau, T. J., Jr., X. Zeng, R. D. Dixon, A. Ouyed, H. Su, and W. Cui, 2023: Tropical mesoscale convective system formation environments. *Atmos. Sci. Lett.*, **24**, e1152, https://doi.org/10.1002/asl.1152.

Gray, W. M., 1998: The formation of tropical cyclones. *Meteor. Atmos. Phys.*, **67**, 37–69, https://doi.org/10.1007/BF01277501.

Haarsma, R. J., and Coauthors, 2016: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geosci. Model Dev.*, **9**, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016.

Hersbach, H., and Coauthors, 2017: Complete ERA5 from 1940: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service (C3S) Data Store (CDS), accessed 30 January 2025, https://doi.org/10.24381/cds.143582cf.

——, and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Holloway, C. E., and J. D. Neelin, 2009: Moisture vertical structure, column water vapor, and tropical deep convection. *J. Atmos. Sci.*, **66**, 1665–1683, https://doi.org/10.1175/2008JAS2806.1.

——, and ——, 2010: Temporal relations of column water vapor and tropical precipitation. *J. Atmos. Sci.*, **67**, 1091–1105, https://doi.org/10.1175/2009JAS3284.1.

Houze, R. A., Jr., 1977: Structure and dynamics of a tropical squall–line system. *Mon. Wea. Rev.*, **105**, 1540–1567, https://doi.org/10.1175/1520-0493(1977)105<1540:SADOAT>2.0.CO;2.

——, 2004: Mesoscale convective systems. *Rev. Geophys.*, **42**, RG4003, https://doi.org/10.1029/2004RG000150.

——, 2018: 100 years of research on mesoscale convective systems. A *Century of Progress in Atmospheric and Related Sciences: Celebrating the American Meteorological Society Centennial*, *Meteor. Monogr.*, No. 59, Amer. Meteor. Soc., https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0001.1.

Huang, X., C. Hu, X. Huang, Y. Chu, Y.-H. Tseng, G. J. Zhang, and Y. Lin, 2018: A long-term tropical mesoscale convective systems dataset based on a novel objective automatic tracking algorithm. *Climate Dyn.*, **51**, 3145–3159, https://doi.org/10.1007/s00382-018-4071-0.

Huffman, G. J., D. T. Bolvin, and E. J. Nelkin, 2015: Integrated Multi-satellitE Retrievals for GPM (IMERG) technical documentation. NASA/GSFC Code 612 Tech. Doc., 48 pp., http://pmm.nasa.gov/sites/default/files/document_files/IMERG_doc.pdf.

Janowiak, J., B. Joyce, and P. Xie, 2023: NCEP/CPC L3 half hourly 4km global (60S - 60N) merged IR V1. Accessed 30 January 2025, https://doi.org/10.5067/P4HZB9N27EKU.

Johnston, B. R., W. J. Randel, and J. P. Sjoberg, 2021: Evaluation of tropospheric moisture characteristics among COSMIC-2, ERA5 and MERRA-2 in the tropics and subtropics. *Remote Sens.*, **13**, 880, https://doi.org/10.3390/rs13050880.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS). *Bull. Amer. Meteor. Soc.*, **91**, 363–376, https://doi.org/10.1175/2009BAMS2755.1.

Kukulies, J., and Coauthors, 2023: Mesoscale convective systems in the third pole region: Characteristics, mechanisms and impact on precipitation. *Front. Earth Sci.*, **11**, 1143380, https://doi.org/10.3389/feart.2023.1143380.

Kuo, H. L., 1974: Further studies of the parameterization of the influence of cumulus convection on large-scale flow. *J. Atmos. Sci.*, **31**, 1232–1240, https://doi.org/10.1175/1520-0469(1974)031<1232:FSOTPO>2.0.CO;2.

Laing, A. G., and J. M. Fritsch, 1997: The global population of mesoscale convective complexes. *Quart. J. Roy. Meteor. Soc.*, **123**, 389–405, https://doi.org/10.1002/qj.49712353807.

——, and ——, 2000: The large-scale environments of the global populations of mesoscale convective complexes. *Mon. Wea. Rev.*, **128**, 2756–2776, https://doi.org/10.1175/1520-0493(2000)128<2756:TLSEOT>2.0.CO;2.

LeMone, M. A., G. M. Barnes, and E. J. Zipser, 1984: Momentum flux by lines of cumulonimbus over the tropical oceans. *J. Atmos. Sci.*, **41**, 1914–1932, https://doi.org/10.1175/1520-0469(1984)041<1914:MFBLOC>2.0.CO;2.

——, E. J. Zipser, and S. B. Trier, 1998: The role of environmental shear and thermodynamic conditions in determining the structure and evolution of mesoscale convective systems during TOGA COARE. *J. Atmos. Sci.*, **55**, 3493–3518, https://doi.org/10.1175/1520-0469(1998)055<3493:TROESA>2.0.CO;2.

Li, J., Y. Li, T. Zhao, R. Schiemann, M. Muetzelfeldt, and X. Jiang, 2021: Northeastward propagation of nocturnal precipitation over the Sichuan Basin. *Int. J. Climatol.*, **41**, E2863–E2879, https://doi.org/10.1002/joc.6886.

Li, P., M. Muetzelfeldt, R. Schiemann, H. Chen, J. Li, K. Furtado, and M. Zhuang, 2023a: Sensitivity of simulated mesoscale convective systems over East Asia to the treatment of convection in a high-resolution GCM. *Climate Dyn.*, **60**, 2783–2801, https://doi.org/10.1007/s00382-022-06471-2.

——, and Coauthors, 2023b: Intensification of mesoscale convective systems in the East Asian rainband over the past two decades. *Geophys. Res. Lett.*, **50**, e2023GL103595, https://doi.org/10.1029/2023GL103595.

Liu, N., L. R. Leung, and Z. Feng, 2021: Global mesoscale convective system latent heating characteristics from GPM retrievals and an MCS tracking dataset. *J. Climate*, **34**, 8599–8613, https://doi.org/10.1175/JCLI-D-20-0997.1.

Maddox, R. A., 1980: Mesoscale convective complexes. *Bull. Amer. Meteor. Soc.*, **61**, 1374–1387.

Mapes, B. E., 1993: Gregarious tropical convection. *J. Atmos. Sci.*, **50**, 2026–2037, https://doi.org/10.1175/1520-0469(1993)050<2026:GTC>2.0.CO;2.

——, 1997: Equilibrium vs. activation control of large-scale variations of tropical deep convection. *The Physics and Parameterization of Moist Atmospheric Convection*, Springer, 321–358, https://doi.org/10.1007/978-94-015-8828-7_13.

——, and R. Neale, 2011: Parameterizing convective organization to escape the entrainment dilemma. *J. Adv. Model. Earth Syst.*, **3**, M06004, https://doi.org/10.1029/2011MS000042.

——, E. S. Chung, W. M. Hannah, H. Masunaga, A. J. Wimmers, and C. S. Velden, 2018: The meandering margin of the meteorological moist tropics. *Geophys. Res. Lett.*, **45**, 1177–1184, https://doi.org/10.1002/2017GL076440.

McAnelly, R. L., and W. R. Cotton, 1986: Meso-β-scale characteristics of an episode of meso-α-scale convective complexes. *Mon. Wea. Rev.*, **114**, 1740–1770, https://doi.org/10.1175/1520-0493(1986)114<1740:MSCOAE>2.0.CO;2.

Mohr, K. I., and E. J. Zipser, 1996: Mesoscale convective systems defined by their 85-GHz ice scattering signature: Size and intensity comparison over tropical oceans and continents. *Mon. Wea. Rev.*, **124**, 2417–2437, https://doi.org/10.1175/1520-0493(1996)124<2417:MCSDBT>2.0.CO;2.

Moncrieff, M. W., and C. Liu, 2006: Representing convective organization in prediction models by a hybrid strategy. *J. Atmos. Sci.*, **63**, 3404–3420, https://doi.org/10.1175/JAS3812.1.

——, ——, and P. Bogenschutz, 2017: Simulation, modeling, and dynamically based parameterization of organized tropical convection for global climate models. *J. Atmos. Sci.*, **74**, 1363–1380, https://doi.org/10.1175/JAS-D-16-0166.1.

Muetzelfeldt, M. R., R. S. Plant, P. A. Clark, A. J. Stirling, and S. J. Woolnough, 2021a: A climatology of tropical wind shear produced by clustering wind profiles from the Met Office Unified Model (GA7.0). *Geosci. Model Dev.*, **14**, 4035–4049, https://doi.org/10.5194/gmd-14-4035-2021.

——, R. Schiemann, A. G. Turner, N. P. Klingaman, P. L. Vidale, and M. J. Roberts, 2021b: Evaluation of Asian summer precipitation in different configurations of a high-resolution general circulation model in a range of decision-relevant spatial scales. *Hydrol. Earth Syst. Sci.*, **25**, 6381–6405, https://doi.org/10.5194/hess-25-6381-2021.

Nesbitt, S. W., R. Cifelli, and S. A. Rutledge, 2006: Storm morphology and rainfall characteristics of TRMM precipitation features. *Mon. Wea. Rev.*, **134**, 2702–2721, https://doi.org/10.1175/MWR3200.1.

Núñez Ocasio, K. M., J. L. Evans, and G. S. Young, 2020: A wave-relative framework analysis of AEW–MCS interactions leading to tropical cyclogenesis. *Mon. Wea. Rev.*, **148**, 4657–4671, https://doi.org/10.1175/MWR-D-20-0152.1.

Park, S., 2014: A Unified Convection scheme (UNICON). Part I: Formulation. *J. Atmos. Sci.*, **71**, 3902–3930, https://doi.org/10.1175/JAS-D-13-0233.1.

Peters, J. M., and R. S. Schumacher, 2014: Objective categorization of heavy-rain-producing MCS synoptic types by rotated principal component analysis. *Mon. Wea. Rev.*, **142**, 1716–1737, https://doi.org/10.1175/MWR-D-13-00295.1.

——, D. R. Chavas, C.-Y. Su, H. Morrison, and B. E. Coffer, 2023: An analytic formula for entraining CAPE in midlatitude storm environments. *J. Atmos. Sci.*, **80**, 2165–2186, https://doi.org/10.1175/JAS-D-23-0003.1.

Peters, O., J. D. Neelin, and S. W. Nesbitt, 2009: Mesoscale convective systems and critical clusters. *J. Atmos. Sci.*, **66**, 2913–2924, https://doi.org/10.1175/2008JAS2761.1.

Rasmussen, K. L., M. M. Chaplin, M. D. Zuluaga, and R. A. Houze Jr., 2016: Contribution of extreme convective storms to rainfall in South America. *J. Hydrometeor.*, **17**, 353–367, https://doi.org/10.1175/JHM-D-15-0067.1.

Rotunno, R., J. B. Klemp, and M. L. Weisman, 1988: A theory for strong, long-lived squall lines. *J. Atmos. Sci.*, **45**, 463–485, https://doi.org/10.1175/1520-0469(1988)045<0463:ATFSLL>2.0.CO;2.

Rutz, J. J., and Coauthors, 2019: The Atmospheric River Tracking Method Intercomparison Project (ARTMIP): Quantifying uncertainties in atmospheric river climatology. *J. Geophys. Res. Atmos.*, **124**, 13 777–13 802, https://doi.org/10.1029/2019JD030936.

Schiro, K. A., S. C. Sullivan, Y.-H. Kuo, H. Su, P. Gentine, G. S. Elsaesser, J. H. Jiang, and J. D. Neelin, 2020: Environmental controls on tropical mesoscale convective system precipitation intensity. *J. Atmos. Sci.*, **77**, 4233–4249, https://doi.org/10.1175/JAS-D-20-0111.1.

Schumacher, C., R. A. Houze Jr., and I. Kraucunas, 2004: The tropical dynamical response to latent heating estimates derived from the TRMM precipitation radar. *J. Atmos. Sci.*, **61**, 1341–1358, https://doi.org/10.1175/1520-0469(2004)061<1341:TTDRTL>2.0.CO;2.

Schumacher, R. S., and K. L. Rasmussen, 2020: The formation, character and changing nature of mesoscale convective systems. *Nat. Rev. Earth Environ.*, **1**, 300–314, https://doi.org/10.1038/s43017-020-0057-7.

Schwendike, J., N. Kalthoff, and M. Kohler, 2010: The impact of mesoscale convective systems on the surface and boundary-layer structure in West Africa: Case-studies from the AMMA campaign 2006. *Quart. J. Roy. Meteor. Soc.*, **136**, 566–582, https://doi.org/10.1002/qj.599.

Simmons, A., and Coauthors, 2020: ERA5.1: Rerun of the fifth generation of ECMWF atmospheric reanalyses of the global climate (2000–2006 only). Copernicus Climate Change Service (C3S) Data Store (CDS), accessed 30 January 2025, https://doi.org/10.24381/cds.143582cf.

Taszarek, M., N. Pilguj, J. T. Allen, V. Gensini, H. E. Brooks, and P. Szuster, 2021: Comparison of convective parameters derived from ERA5 and MERRA-2 with rawinsonde data over Europe and North America. *J. Climate*, **34**, 3211–3237, https://doi.org/10.1175/JCLI-D-20-0484.1.

Thorpe, A. J., M. J. Miller, and M. W. Moncrieff, 1982: Two-dimensional convection in non-constant shear: A model of mid-latitude squall lines. *Quart. J. Roy. Meteor. Soc.*, **108**, 739–762, https://doi.org/10.1002/qj.49710845802.

Virman, M., M. Bister, J. Räisänen, V. A. Sinclair, and H. Järvinen, 2021: Radiosonde comparison of ERA5 and ERA-Interim reanalysis datasets over tropical oceans. *Tellus*, **73A**, 1929752, https://doi.org/10.1080/16000870.2021.1929752.

Wall, C. J., D. L. Hartmann, M. M. Thieman, W. L. Smith Jr., and P. Minnis, 2018: The life cycle of anvil clouds and the top-of-atmosphere radiation balance over the tropical West Pacific. *J. Climate*, **31**, 10 059–10 080, https://doi.org/10.1175/JCLI-D-18-0154.1.

Yano, J.-I., and R. S. Plant, 2012: Convective quasi-equilibrium. *Rev. Geophys.*, **50**, RG4004, https://doi.org/10.1029/2011RG000378.

Yuan, J., and R. A. Houze Jr., 2010: Global variability of mesoscale convective system anvil structure from A-Train satellite data. *J. Climate*, **23**, 5864–5888, https://doi.org/10.1175/2010JCLI3671.1.

Yun, Y., C. Liu, Y. Luo, and W. Gao, 2021: Warm-season mesoscale convective systems over eastern China: Convection-permitting climate model simulation and observation. *Climate Dyn.*, **57**, 3599–3617, https://doi.org/10.1007/s00382-021-05994-4.

Zhang, C., B. E. Mapes, and B. J. Soden, 2003: Bimodality in tropical water vapour. *Quart. J. Roy. Meteor. Soc.*, **129**, 2847–2866, https://doi.org/10.1256/qj.02.166.

Zipser, E. J., 1977: Mesoscale and convective–scale downdrafts as distinct components of squall-line structure. *Mon. Wea. Rev.*, **105**, 1568–1589, https://doi.org/10.1175/1520-0493(1977)105<1568:MACDAD>2.0.CO;2.