Tropical Oceanic Mesoscale Cold Pools in a High-Resolution Global Cloud-Resolving Model from DYAMOND Initiative

Piyush Garg[1*], Stephen W. Nesbitt[1], Timothy J. Lang[2], Georgios Priftis[3]

[1]Department of Atmospheric Sciences, University of Illinois Urbana-Champaign, Urbana, IL USA
[2]NASA Marshall Space Flight Center, Huntsville, AL USA
[3]University of Alabama-Huntsville, Huntsville, AL USA

Key Points:

- A 40-day 2.5 km global cloud-resolving model simulation is used to identify and characterize tropical oceanic mesoscale cold pools.
- Model-simulated cold pool frequency, size, and precipitation compare well with scatterometer-identified cold pools.
- Random forest regression is applied to identify environmental controls on simulated cold pool frequency, size, and intensity.

*Current Affiliation: Argonne National Laboratory, Lemont, IL

Corresponding author: Piyush Garg, pgarg7@illinois.edu

Abstract

In recent years, global kilometer-scale convection-permitting models have shown promising results in producing realistic convection and precipitation. Cold pools, which can be represented by km-scale models, are identified as one of the significant mesoscale processes responsible for modulating the life cycle of mesoscale organized convection. However, there is still a lack of understanding about cold pool properties across the spatio-temporal scales, as well as their representation in models. In this study, a 2.5 km global Icosahedral Nonhydrostatic (ICON) model simulation run for 40 days (06 UTC 01 Aug - 23 UTC 10 Sep 2016) from the Dynamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains (DYAMOND) initiative is used to identify thermal cold pools (using T_v) over the tropical oceans. The diurnal cycle of simulated thermal cold pools is compared against NASA's RapidScat-observed gradient feature (GF) frequency and IMERG precipitation. ICON and IMERG exhibit morning peaks in cold pool activity similar to RapidScat GF frequency but miss the RapidScat-observed afternoon peak. EUMETSAT's Advanced Scatterometer (ASCAT) and RapidScat GF spatial climatology is also compared to ICON cold pools, where ICON shows more cold pools over the Indo-Pacific and western Atlantic basins. ICON TF size and precipitation percentiles are validated with ASCAT and RapidScat GF size and precipitation, and the simulated cold pool properties depict similar regional variability in cold pool properties with a smaller order of magnitude. Random forest regression is applied to identify critical environmental properties with column water vapor found out to be most important for controlling cold pool number, size, and intensity. Regional differences between cold pool properties are explored, where easterly waves dominate the eastern Pacific and Atlantic cold pool activity. The western Pacific and the Indian Ocean cold pools are controlled by local mesoscale forcing and intraseasonal oscillations. Thus, a holistic conceptual framework is established to explain the simulated cold pool characteristics over tropical oceans.

Plain Language Summary

Cold pools are the cooler-than-ambient temperature regions formed when precipitation reaches near-surface and cools the air in the vicinity. Their boundaries act as density currents due to the difference in the ambient environment and in-pool air density. Air can lift on these boundaries to initiate secondary convection, and thus these cold pools are important for moist convection. There is still a lack of understanding about the relationship between cold pools and their ambient environment. Storm-resolving models are an excellent tool to analyze cold pools in a near-realistic atmospheric state, thus motivating this study. The first objective is to identify thermal cold pools in a global high-resolution storm-resolving model and validate against satellite-observed cold pool climatology. Understanding biases in simulated cold pool properties is essential to improve the model physics further. Secondly, a thorough analysis of cold pools and their environmental properties such as total moisture and vertical wind shear is carried out using novel machine learning methods to illustrate better how cold pools relate to their storm environment in a high-resolution model. Lastly, a conceptual understanding is established to explain the controls on cold pool activity over different oceanic basins.

1 Introduction

Global cloud-resolving models (GCRMs) are a recent advance in modeling that can resolve non-hydrostatic accelerations over the global domain with kilometer-scale resolution (Satoh et al., 2019; Stevens et al., 2019). In other words, CRMs use non-hydrostatic equation numerics, which can permit convective cloud buoyancy. In reality, CRMs allow the mesoscale dynamics of precipitating storm systems and are not necessarily able to well-resolve convective updrafts and downdrafts, and thus are also termed as storm-resolving models or SRMs (Guichard & Couvreux, 2017; Satoh et al., 2019). These SRM simulations can be an excellent tool to interpret observations, especially over the oceanic regions where *in situ* observations are sparse in nature (Guichard & Couvreux, 2017), despite under-resolving convective motions. A first-ever intercomparison of a range of GSRMs such as Icosahedral Nonhydrostatic (ICON), Model for Prediction Across Scales (MPAS), and System for Atmospheric Modeling (SAM) was carried out as a part of the Dynamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains (DYAMOND) initiative as explained in Stevens et al. (2019). This intercomparison initiative is one-of-its-kind and provides an opportunity to test a range of hypotheses related to deep convective dynamics. Although these simulations can act as a bridge between observations and parameterized models, such high-resolution models need to be validated against independent observations to understand the bias and uncertainty in resolving transient convection-precipitation processes (Stevens et al., 2019).

Cold pools are envelopes of air produced by precipitating downdrafts, which upon reaching the surface, spreads out and creates a gust front boundary (Simpson, 1969; Simpson et al., 1977). Due to the difference between the density of ambient environmental air and in-pool air, a cold pool acts as a density current. Warmer environmental air can lift on the boundary of these density currents and can initiate secondary convection (Charba, 1974; Wakimoto, 1982; Kingsmill, 1995; Knupp, 2006). During the Global Atmospheric Research Program Atlantic Tropical Experiments (GATE) in the early 1970s, R. A. Houze and Betts (1981) observed that cold pools are integral in modulating air-sea exchanges over the tropical oceans. Since then, an increased curiosity in observing and characterizing cold pools with respect to tropical cloud and precipitation mesoscale organization has been observed (Mapes et al., 2006; Tao et al., 2007; Holloway & Neelin, 2009; Nasuno et al., 2009; Feng et al., 2015; Rowe & Houze, 2015; Kilpatrick & Xie, 2015; Ruppert & Johnson, 2016; de Szoeke et al., 2017; Garg et al., 2020; Cheng et al., 2020; Garg et al., 2021). Mesoscale cold pools have not been appropriately resolved in numerical weather prediction models in the past (Olson, 1985; Stensrud & Fritsch, 1994; Spencer & Stensrud, 1998). Cold pools have also been observed to have a strong relationship with the convective organization, heavy precipitation, and flash flood events (Maddox et al., 1979) and thus including information about cold pool strength, density and location improved heavy precipitation prediction in the models. Numerical model studies such as Cortinas and Stensrud (1995); Trapp and Woznicki (2017) and (Borque et al., 2020), found out that cold pools alter the environmental properties (e.g, CAPE, surface fluxes) and thus an accurate knowledge of the size, intensity, and frequency of cold pools is important for robust model prediction of precipitation. Over tropical oceans, cold pools have been hypothesized to play a crucial role within shallow-to-deep transitions and also in modulating surface energy fluxes (Feng et al., 2015; Chandra et al., 2018; Pei et al., 2018). Also, accurate characterization of cold pools within the tropical environment is essential to correct the erroneous diurnal cycle of precipitation and convection in global climate models (Schlemmer & Hohenegger, 2014; Pei et al., 2018). Therefore, to guide the models to predict the convection-precipitation relationship better, thorough comparisons with available observations need to be carried out, thus leading to the motivation behind this study.

Long-term satellite observations have allowed us to observe the atmosphere-ocean interactions at a range of spatio-temporal resolutions. Space-borne scatterometers (e.g., QuikScat, ASCAT, RapidScat) have provided ocean vector winds across the global oceans and have been used to identify oceanic cold pools utilizing a range of metrics such as horizontal wind divergence (e.g., Mapes et al., 2009; Kilpatrick & Xie, 2015) and wind gradient (Garg et al., 2020, 2021). RapidScat, in particular, was in a non-sun-synchronous orbit onboard the International Space Station (ISS) from 2014-2016 and provided diurnally resolved observations of ocean vector winds. This dataset allows the validation of model-simulated cold pool properties within the diurnal range. RapidScat-identified GF dataset can be an efficient tool for understanding bias in simulated cold pool characteristics regarding density, size, intensity, and the diurnal cycle. Since precipitation, convection, and cold pool activity are tightly linked, analyzing these properties together will help understand their relationships. Tropical Rainfall Measurement Mission (TRMM) and its successor Global Precipitation Measurement (GPM) have been providing near-real-time precipitation data across the globe since 1998. The Integrated Multi-satellite Retrievals for GPM (IMERG) combined precipitation information from the entire GPM satellite constellation and is an instrumental dataset over the observation-sparse regions (e.g., tropical oceans). IMERG provides retrievals of the diurnal cycle of precipitation. Thus, comparing it with cold pool activity from observations and models will prove to be highly beneficial for both the remote sensing and modeling community.

Convective parameterizations have been an integral part of global climate models (GCMs) since Manabe et al. (1965), but they still suffer from uncertainties in resolving precipitation processes. Convection in most GCMs is activated through a trigger function within these convective parameterization schemes. Due to a gap in understanding of convective processes on a subgrid-scale, unrealistic simulations of the diurnal cycle of convection, Madden-Julian Oscillation (MJO), and the intertropical convergence zone (ITCZ) have been observed in previous studies (e.g., Xie et al., 2004; Lin et al., 2008; Liu et al., 2010). Recently, CRMs have been identified as an efficient tool to identify statistical relationships between environment and convective processes (e.g., convective initiation, MCS life cycle) using a range of machine learning (ML) methods. Once these relationships are identified, GCM convective parameterizations are replaced by ML-based statistical models, and an improvement in resolving precipitation and cloud dynamic processes has been observed (Brenowitz & Bretherton, 2018; Gentine et al., 2018; Rasp et al., 2018; O'Gorman & Dwyer, 2018; Ukkonen & Mäkelä, 2019). Regression and classification methods within the ML framework thus can be successfully implemented to learn how environmental properties (e.g., CAPE. wind shear, relative humidity) affect cold pool properties (e.g., number and intensity) from a CRM simulation. This exercise can result in obtaining information about which features have a higher weightage in producing cold pools and thus will be extremely useful to improve cold pool parameterization in GCMs. The difference in convective dynamics over different oceanic basins can cause different relationships between the ambient environment and cold pools. Thus basin-specific understanding needs to be established to create more physically coherent and robust model architectures.

The objectives of this study are twofold. First, global climatological properties of cold pools are analyzed from the 40-day global high-resolution ICON model simulation obtained from the DYAMOND protocol and are compared with ASCAT and RapidScat cold pool climatology from Garg et al. (2020) and Garg et al. (2021). Cold pool size and precipitation percentiles are also compared with ASCAT- and RapidScat-identified cold pools to gauge how well ICON can produce cold pools compared to climatological variability. The diurnal cycle of ICON-simulated cold pools is compared to RapidScat-identified cold pool and IMERG-observed precipitation diurnal cycle. In the second objective, once an observation-model comparison is obtained, ML regression is applied to identify the importance of features relevant to cold pool number density, size, and intensity. Regional differences in cold pool properties are identified, and possible physical mechanisms are explored. This study is organized as follows. Section 2 covers the datasets and methodology used in this analysis. Section 3 depicts the global climatologies of cold pools and their attributed environmental properties. Section 3 also shows the comparison between ICON, ASCAT, and RapidScat-identified cold pools. Section 4 explores the diurnal cycle of ICON-simulated and RapidScat-observed cold pools with IMERG precipitation. Section 5 shows the application of ML regression to identify important environmental parameters for cold pool dynamics. Section 6 explores the regional differences in the cold pool-environment relationship. Section 7 summarizes the results and concludes the study.

2 Data and Methodology

2.1 Satellite Datasets

NASA's RapidScat was a Ku-band (13.4 GHz), conically scanning two-beam spaceborne scatterometer onboard the ISS from 03 October 2014 - 19 August 2016 in a non-sun-synchronous orbit. RapidScat retrieved 10 m ocean vector winds at 12.5 km field of view (FOV) during the two years of the operational period. The ground swath width was 900 km with an incidence angle of 49^o and a slant range of 600 km. The satellite operated at 92.5% uptime with instrument outages related to ISS vehicular docking (Lang, 2017). Most of the reductions in the observational quality of the data was due to change in the altitude and attitude of ISS. This study uses level 2B 12.5 km FOV version 1.0 climate quality ocean wind vectors provided by NASA - Physical Oceanography Distributed Active Archive Center (NASA PODAAC; SeaPAC, 2015) for the entire operational period of RapidScat. Cold pools are identified by applying the gradient feature (GF) algorithm from Garg et al. (2020) on the RapidScat vector wind data, as depicted in Garg et al. (2021). Advanced Scatterometer (ASCAT) was in a sun-synchronous orbit onboard European organization for the Exploitation of Meteorological Satellites (EUMETSAT) Meteorological Operational - A (MetOp-A) satellite from 2007-2021 and provided 10 m surface winds at 12.5 km FOV. ASCAT has two swaths of 500 km separated by a distance of 360 km in between and crosses geolocation twice a day (approx. 9 AM and 9 PM local time). Due to a longer temporal record, this study also uses Advanced Scatterometer (ASCAT) GFs from Garg et al. (2020) from 01 Aug - 10 Sep for 2007-2018 to compare ICON TF frequency with GF frequency climatology.

IMERG precipitation data is a globally merged dataset consisting of all satellite microwave precipitation estimates with microwave calibrated infrared (IR) satellite estimates and precipitation gauge analyses for the TRMM and GPM time period across the globe. This study uses 30-min, 0.1^o, IMERG precipitation data from 2000-2020 obtained from NASA Goddard Earth Sciences Data and Information Services Center (GES DISC; Huffman et al., 2019). Note that all land area is masked to perform the precipitation analysis in this study. All the data from RapidScat and IMERG is converted from UTC to local time (LT) to perform the diurnal analysis in this study.

2.2 ICON Model Simulation

The DYAMOND experimental protocol ran a simulation of 40 days and 40 nights from 00 UTC 1 August 2016 with a range of GSRMs at a grid spacing of 5 km or less. The protocol used the initialization date of 1 August 2016 to link the DYAMOND runs with previous large-domain SRM simulation runs supporting field campaigns over the Northern Atlantic (Klocke et al., 2017; Stevens et al., 2019). This 40-day period also coincided with the northern hemispheric monsoon and an active tropical cyclone season. All the modeling groups applied a hierarchical approach of frequent 2D outputs and relatively less frequent 3D outputs. The simulations produced many MCSs, thus allowing a detailed characterization of the mesoscale convective organization over both land and ocean. All the output fields and postprocessing framework were provided by the German Climate Computing Center (DKRZ; Stevens et al., 2019).

ICON model uses an unstructured triangular grid based on the successive refinement of a spherical icosahedron in which 20 equilateral triangles are present, each of equal size (Zängl et al., 2015). The grid spacing in the ICON model corresponds to the square root of the mean cell area of the model triangles (Hohenegger et al., 2020). The dynamical set of equations in ICON is based on Gassmann and Herzog (2008) in which local mass, energy, and Ertel's potential vorticity (EPV) conservation are achieved by using formulations in the turbulence-averaged form of the relevant turbulent fluxes, radiation fluxes, and by describing a model atmosphere of dry air and water in gaseous, liquid, and solid form. ICON simulation within the DYAMOND protocol follows Hohenegger et al. (2020) where convection and gravity wave drag parameterizations were not used, and graupel was utilized as an additional prognostic variable in the bulk microphysics scheme consisting of rain and cloud water, cloud ice, and snow as well. Thus, the ICON simulations used in this study were convection-permitting for both shallow and deep convection. This study uses R2B10 (2.5 km) resolution for the analysis performed here out of all the available grid spacings. The temporal resolution of 2D and 3D fields was 15 minutes and 3 hours, respectively. The data was regridded from icosahedral grid to regular lat-lon grid before using it for the analysis using climate data operators (CDO; Schulzweida et al., 2006) tool provided by DKRZ on the Mistral supercomputing cluster. In the vertical, ICON used 90 levels with the model top at 75 km with damping beginning in the 77$^{\text{th}}$ layer, above 44 km. The 40-day simulation was initialized using European Center for Medium-Range Weather Forecasts (ECMWF) runs at 00 UTC 01 August 2016. Prescribed SST and sea ice cover were used for the initialization. The model time step was 22.5 s for 2.5 km resolution. Data for the bottom boundary condition was recreated by aggregating climatological mean near-surface temperature, aerosol optical depth, soil albedos, soil texture, normalized differential vegetation index, and remaining albedo values. Table 1 summarizes the physical parameterizations used for ICON simulation within the DYAMOND protocol as mentioned in Hohenegger et al. (2020). These schemes are derived from the Consortium for Small-scale Modeling (COSMO) model, which has been used to run simulations at smaller resolutions (Fuhrer et al., 2018). Cold pools in the ICON simulation are identified using virtual temperature anomaly (T_v), similar to Garg et al. (2020) WRF-based cold pool identification using equation 1 below. A *Gaussian* filter with σ of 100 was applied on the raw (T_v) field calculated using equation 1. To be consistent with the GF terminology in Garg et al. (2020), T_v-identified cold pools are called temperature features or TFs throughout this study.

$$T_v = (1 + 0.61q)T \tag{1}$$

Where q and T are 2m specific humidity and air temperature, respectively.

3 Global Tropical Oceanic Cold Pool Climatology

3.1 ICON-Simulated Cold Pools and their Characteristics

Scatterometer-observed cold pool number density, size, and precipitation climatology in Garg et al. (2020) and Garg et al. (2021) match well with the global MCS distribution analyzed in previous studies (Nesbitt et al., 2006; Houze et al., 2015; Huang et al., 2018; Feng et al., 2021). ASCAT- and RapidScat-observed cold pool properties are compared with ICON-simulated cold pool spatial climatology. All 40-day ICON-simulated cold pool mean climatologies in Fig. 1-3 are calculated in a 0.5° grid box. Fig. 1a shows the western Pacific (WPAC), eastern Pacific (EPAC), Atlantic (AO), and Indian (IO) ocean basin analysis regions similar to Garg et al. (2021). Table 2 shows the latitude and longitude of all four region edges. All values below 10$^{\text{th}}$ percentile of TF number density in a grid box are removed from the analyses to only observe the mean values of cold pool-attributed properties which do not suffer from low count bias. In Fig. 1a, the highest TF density is within the northern Pacific and Atlantic ITCZ (3500 and above), followed by the western Pacific warm pool region (3000 - 3800), the SPCZ (2000 - 3000) and the Indian Ocean (1500 - 3000). These numbers correspond well spatially with the regions of active, organized deep convection (Nesbitt et al., 2006; Houze et al., 2015; Feng et al., 2021). Comparing trade wind regions with ITCZ, especially the southeastern and northeastern Pacific and Atlantic, have negligible TF activity compared to the ITCZ and SPCZ. This number density signature matches well with the ASCAT- and RapidScat-observed GF climatology in Garg et al. (2020) and Garg et al. (2021), where most of the mesoscale GFs were within the ITCZ and SPCZ. Note that since this is only a 40-day climatology, the seasonal migration of ITCZ will not be visible in the results here as compared to satellite-observed GF climatology, where GFs were visible in both Northern and Southern Hemisphere ITCZ.

Figure 1b shows another cold pool parameter; TF size (km^2). Mean largest TFs are observed over the Bay of Bengal ($\geq$ 200 km^2), northern Indian Ocean (175 - 200 km^2), Maritime Continent (175 - 200 km^2), and western Pacific tropical cyclone region (150 - 175 km^2). Interestingly, all other regions over the tropical oceans have TFs between 50 - 125 km^2, except the Atlantic ITCZ having relatively larger GFs (150 - 190 km^2) than the rest of the Atlantic basin. Mean larger TFs in the model simulation can be attributed to the active boreal monsoonal circulation over the Bay of Bengal, the Indian Ocean, and the Maritime Continent. Also, ICON simulated a number of tropical cyclones over the Bay of Bengal, Maritime Continent, and the western Pacific (Stevens et al., 2019), which could have led to intense convective activity in the region, thus resulting into larger cold pools within the outer rain bands of tropical cyclones and MCSs. TF size in Fig. 1b differs from ASCAT- and RapidScat-observed GF size climatology (Garg et al., 2020, 2021) as the largest GFs were observed over eastern-central Pacific, followed by WPAC, Indian Ocean and Atlantic Ocean in the scatterometer-observed cold pool climatology. This difference between the simulated TF and observed GF size climatology could be due to the simulation not being able to capture intra-seasonal to interannual oscillations, as it is only a 40-day simulation. However, TF size makes physical sense in terms of organized convective activity as shown in Stevens et al. (2019) where the authors observed vigorous convection over Indo-Pacific warm pool and Maritime Continent.

Mean precipitation within the TF polygons is shown in Fig. 1c. The eastern Pacific basin shows the heaviest TF precipitation ($\geq$ 12.5 mm hr^{-1}), followed by coastal regions across the tropical oceans (10 - 17 mm hr^{-1}). Heavier TF precipitation over the eastern Pacific could be due to longer-lasting convective clusters and synoptic wave patterns leading to organized convection (Hohenegger et al., 2020). Note the enhancement in TF-attributed precipitation along the coasts of the Bay of Bengal, Maritime Continent, Papua New Guinea, and North and South Americas. This enhanced precipitation on coastlines is likely associated with land-sea breeze circulations observed in previous studies (Yang & Slingo, 2001; Mori et al., 2004; Tang et al., 2019). Other regions of the tropical oceans observe moderately heavy precipitation (5 - 10 mm hr^{-1}) for all the TFs simulated during the entire period.

Total column water vapor (TCWV) has been identified as one of the critical environmental parameters responsible for modulating MCS strength and life cycle (Bretherton et al., 2004; Holloway & Neelin, 2009; Schiro et al., 2016; Schiro & Neelin, 2019). Fig. 1d shows the simulated maximum TCWV corresponding to TFs. Within the ITCZ and SPCZ, the mean maximum TCWV values are above 55 kg m^{-2} which is close to what was observed in Schiro and Neelin (2019) and Garg et al. (2021). TCWV maxima are observed over the Maritime Continent, Bay of Bengal, Arabian Sea, and western Pacific tropical cyclone region. Note that these are the regions of larger TFs observed in Fig. 1b. Overall, the TCWV values depict that TCWV is relatively similar when compared between different basins within the regions of deep tropical convection. The key takeaway points from Fig. 1 are that (a) number of TFs are higher in the region of frequent deep convection, (b) TF size is relatively independent of TF number density as in Garg et al. (2020) and Garg et al. (2021), (c) heavier precipitation does not necessarily result in numerous or larger cold pools, (d) water vapor is a critical parameter in modulating cold pool number and size.

TFs in ICON are identified using T_v anomaly. In order to identify where the model produces the most intense cold pools, the mean T_v anomaly for all TFs in a grid box is shown in Fig. 2a. Also, horizontal wind gradient has been used as an identification parameter in Garg et al. (2020, 2021). Therefore, looking at T_v anomaly and wind gradient (Fig. 2b) together would provide us an idea about the location of colder (warmer) cold pools with stronger (weaker) wind gradients. Comparing Fig. 2a-b, it can be seen that the coldest TFs ($\leq$ -1.85 K) and strongest wind gradient ($\geq 1 \times 10^{-3}$ s^{-1}) are over the western Pacific ITCZ and tropical cyclogenesis region, SPCZ, Indian Ocean ITCZ, Arabian Sea, Bay of Bengal, and northwestern Atlantic near the Gulf of Mexico, and the Caribbean Sea. Note that these are the regions of high TF number density observed in Fig. 1a, which makes sense as colder cold pools would produce stronger wind gradients at their boundary (Wills et al., 2021) and result in secondary cold pool formation in the vicinity, thus increasing the number of TFs in these regions. Good correspondence between T_v anomaly and wind gradient provides further validation to the GF hypothesis (Garg et al., 2020) as well. Fig. 2c shows relative humidity (RH) within the TFs. The primary aim to look at RH in addition to TCWV is that TCWV is a measure of total gaseous water contained in a vertical column in the atmosphere, while RH is the net amount of water vapor in the air relative to the amount of water vapor the air is capable of holding. Therefore, RH can provide information about the available amount of water within the cold pool rather than the total water vapor. Hence, RH (Fig. 2c) depicts a different signal than TCWV (Fig. 1d), where the highest RH (≥ 0.94) is over the eastern Pacific and Atlantic Ocean basins while TCWV maxima are over the Maritime Continent and the Indian Ocean. However, over the region of intense convection and cold pool activity, RH is over 0.9 in general, which is similar in principle to TCWV, where deep tropics have consistent values of $\sim$ 55 kg m^{-2} across different basins.

CAPE is generally defined as the vertically integrated buoyancy of adiabatically lifted sub-cloud air and has been a good predictor of thunderstorm severity (Brooks et al., 1994), lightning flash rates (Williams et al., 1992), and precipitation extremes (Lepore et al., 2015, 2018). CAPE is also used in most convective parameterizations in contemporary GCMs to compute the cloud base mass flux, which is responsible for controlling the convective heating and coverage in climate scale simulations (Lin & Neelin, 2003; Diffenbaugh et al., 2013; Romps et al., 2014; Seeley & Romps, 2015). Since cold pools affect the buoyancy of environment air parcels, CAPE is an important parameter for analyzing cold pool properties across the global tropics. Fig. 2d shows the mean maximum CAPE within a 2^{o} buffer of TF polygons since maximum CAPE does not necessarily exist within the cold pool. In this way, maximum CAPE is identified within the respective TF regions with an added buffer. High values of CAPE are observed over the Gulf of Mexico near the coastal United States and the Caribbean Sea ($\geq$ 1000 J kg^{-1}), followed by western Pacific ITCZ and tropical cyclogenesis regions, SPCZ, and Indian Ocean ITCZ (800 - 1000 J kg^{-1}). Over the rest of the deep tropics, the CAPE values are in the range of 400 and 600 J kg^{-1}. Bhat et al. (1996) used datasets from tropical sounding sites and observed that CAPE, which depends on the surface and upper-air thermodynamic properties, is an integral quantity to link deep moist convection with surface properties over tropical oceans. Comparing Fig. 2d with Fig. 2a-b, it can be seen that a maximum in CAPE coincides with colder T_v anomaly and a higher wind gradient, thus suggesting that a higher CAPE should result in more intense cold pools at the surface. The key points from Fig. 2 are that (a) colder cold pools have stronger wind gradients at their boundary and coincide with the number density of TFs, (b) most of the deep tropics have cold pool humidity above 90%, but the RH maxima have a different location than TCWV maxima, and (c) CAPE is an important environmental parameter crucial for defining cold pool intensity at the surface.

Cold pools have been observed to modify surface sensible (SHF) and latent heat (LHF) fluxes in a range of observational and model-based studies (e.g., Feng et al., 2015; de Szoeke et al., 2017; Zuidema et al., 2017). In the bulk aerodynamic formulae, SHF is a function of wind speeds and the difference between air and sea surface temperature while LHF is a function of wind speeds and the difference between atmospheric and sea surface humidity (Vickers & Mahrt, 2006). Since cold pools affect temperature, humidity, and wind speeds, surface fluxes are important in understanding atmosphere-ocean exchange processes during and after the passage of cold pools. Fig. 3a-b shows the mean SHF and LHF within TFs for the entire simulation. Comparing Figs. 3a-b, it can be seen that there are regions in the southern hemisphere (the Indian Ocean, near Australian coastal regions, southern Pacific, and the Atlantic Ocean), where both SHF and LHF show similar signatures, which could mean that these areas are dominated by the change in wind speeds due to cold pool passage as both the fluxes have wind speed effect in common. In Fig. 3a, most of the deep tropics have moderate SHF between 40 - 70 Wm^{-2} with pockets of maxima over the Bay of Bengal, Maritime Continent, and coastal United States (values around 75 - 80 Wm^{-2}). These local maxima signify an enhanced temperature gradient between the sea surface and atmosphere due to cold pool activity and coincide with TF density (Fig. 2a), size (Fig. 1b), and CAPE (Fig. 2d). Fig. 3b shows the LHF maxima over the Bay of Bengal, Maritime Continent, western Pacific tropical cyclone region ($\geq$ 250 Wm^{-2}), followed by the eastern and western coast of the United States and Caribbean Sea (200 - 250 Wm^{-2}). All other tropical oceanic regions have LHF values in the range of 100 - 200 Wm^{-2}. LHF maxima coincide with TF size (Fig. 1b) and TCWV (Fig. 1d) maxima, in particular, suggesting that humidity gradient between the atmosphere and the ocean play a crucial role in modulating the area of cold pools in these regions and vice versa.

Vertical wind shear plays a critical role in organizing atmospheric moist convection into a variety of systems ranging from supercells to tropical cloud clusters and squall lines (Cotton et al., 1995; Houze et al., 1993). Rotunno et al. (1988) and Weisman et al. (1988) argued that cold pool - wind shear interaction could prolong the lifetime of squall lines and influence their intensity. It has also been observed that when the cold pool is roughly balanced with wind shear (also known as the optimal state), the system can maintain an upright updraft and can initiate secondary convection on the leading edge of the cold pool boundary (Xu & Randall, 1996; Xue, 2000; Robe & Emanuel, 2001). Fig. 3c-d shows the vertical wind shear between surface - 600 hPa and surface - 300 hPa, respectively. The aim of looking at these wind shear profiles is to understand the relationship of mid-tropospheric and upper-tropospheric wind shear with TF properties. In both Fig. 3c-d, the Southern hemisphere (south of 20°S) has highest vertical wind shear; 15 - 20 ms^{-1} and $\geq$ 30 ms^{-1} for the mid-and upper-troposphere respectively. The high wind shear values in surface - 300 hPa (Fig. 3d) signify the seasonal mean location of the subtropical jet stream in the southern hemisphere (e.g., Gallego et al., 2005). Apart from this southern hemispheric region, the mid-tropospheric wind shear (Fig. 3c) is between 5 - 10 ms^{-1} and upper-tropospheric wind shear (Fig. 3d) is between 5 - 20 ms^{-1} across the tropical oceanic basins. Mid-tropospheric wind shear (Fig. 3c) shows relatively higher values near the equator than trade wind regions with maxima over the eastern Pacific followed by the Bay of Bengal, Arabian Sea, and the Atlantic Ocean. Note the relatively strong upper-tropospheric wind shear (Fig. 3d) over the Indian Ocean and Maritime Continent shows a signal of the tropical easterly jet (TEJ), which is commonly observed during the Indian summer monsoon (ISM) season (Koteswaram & George, 1958; Roja Raman et al., 2009; Huang et al., 2021). Also, the northeastern Pacific and Atlantic Ocean (north of 20°N) show enhanced wind shear in surface - 300 hPa layer (Fig. 3d), suggesting the presence of the subtropical jet stream over these tropical regions (e.g., Nakamura et al., 2004). Overall, the vertical wind shear profiles examined in these two layers showed that most cold pools within the regions of deep moist convection over tropical oceans have low-to-moderate vertical wind shear for both surface to 600 and 300 hPa. Comparing wind shear (Fig. 3c-d) with TF size (Fig. 1b) and TCWV (Fig. 1d), it can be said that the relatively high wind shear over the Indian Ocean and Maritime Continent corresponds well with larger and relatively moister TFs. The key takeaway points from Fig. 3 are (a) SHF and LHF maxima signify a larger variation in winds, humidity, and temperature gradient between the ocean and atmosphere and exists in the region of larger and moister TFs with higher CAPE, and (b) relatively moderate-to-high vertical wind shear between the surface to mid-and upper-troposphere relates well with larger and relatively moister TFs.

3.2 Comparison between ICON TFs and RapidScat and ASCAT GFs

Although the ICON TF climatology is physically self-consistent across the global tropics, validation with observations would further provide a comprehensive outlook about the TF properties. In Figure 4, ASCAT's 12-year (2007-2018) and RapidScat's 2-year (2014-2016) GF climatology for 01 Aug - 10 Sep is used to compare with ICON TFs 40-day climatology. Note that only TFs greater than the smallest GFs (300 km^2) are being used to compare the two datasets spatially. Both the datasets are gridded in a 5° grid-box and are normalized by the GF and TF maximum frequency. Figure 4 shows the percentage difference between ASCAT, RapidScat GFs and ICON TFs variation from their respective maximum occurrence frequency, where blues represent higher TF frequency while reds represent more GFs in the grid box. Overall, Fig. 4 suggests that more GFs are observed across most of the global tropics except the Bay of Bengal, Indo-Pacific region, Maritime Continent, west-central Pacific, and western Atlantic. Comparing Fig. 4 with Fig. 1a-b, it can be said that ICON simulation during the Northern Hemisphere summer resulted in frequent mesoscale convection due to the passage of intraseasonal oscillations, thus producing numerous, larger thermal cold pools over the northwestern Pacific, Bay of Bengal, and Maritime Continent. More TFs in the western Atlantic can be due to a number of tropical cyclones produced in the simulation where cold pools may have been produced in the outermost rainbands of TCs, resulting in higher TF frequency (Stevens et al., 2019). Also, these differences could be due to the temporal differences and biases in the model in representing cold pools. Comparing Fig. 4a and 4b, note that the difference in RapidScat and ICON (Fig. 4b) is relatively less than ASCAT and ICON in Fig. 4a. This relatively less difference in GF and TF frequency in Fig. 4b could be due to the fact that RapidScat was able to sample diurnally and thus was able to observe cold pools of different intensities as compared to ASCAT which could only observe cold pools twice a day. Another reason could be that RapidScat had a continuous swath while ASCAT had a gap in between two swaths and thus RapidScat would miss less number of GFs as compared to ASCAT. This observation-model comparison is important in understanding biases in model simulation, which would help further to improve the representation of convective processes in these SRMs. Overall, the comparison suggests that both climatologies do not deviate significantly from each other, thus providing confidence in using ICON simulation to observe global tropical oceanic cold pool characteristics.

The relative frequency of ICON TFs with ASCAT and RapidScat GFs is shown in Fig. 5, which is calculated by dividing TFs (GFs) frequency within each basin with the total TFs (GFs). In the case of WPAC, ICON has relatively more TFs than RapidScat, followed by ASCAT. This observation makes sense as ICON's western Pacific warm pool was quite active during the simulation period, resulting in more convective systems and associated cold pools than the climatological number of GFs from either scatterometer. For EPAC, RapidScat's climatological frequency leads ICON TFs and ASCAT GFs. Stronger convective activity over the eastern Pacific is governed by a range of factors such as the El Niño Southern Oscillation (ENSO) phase and tropical wave activity (D. J. Raymond et al., 2003). During 2014-2016, ENSO was in the positive phase with significant warming over the equatorial eastern Pacific (L'Heureux et al., 2017) and RapidScat's period coincides with this warm anomaly, which might have led to stronger convective events over EPAC and thus more cold pools. ICON's simulation period was in August-September 2016 when ENSO phase transitioned to a negative phase and thus relatively less number of TFs compared to RapidScat GFs. For both WPAC and EPAC, ASCAT GFs are relatively lower in number as compared to rest of the two datasets, thus indicating that a longer period (2007-2018) may have led to a more mean "climatological" cold pool frequency in this analysis. However, ASCAT GFs lead in case of the AO, closely followed by RapidScat and then ICON. The hypothesis remains consistent in this case as a longer temporal observations would provide a more "mean" value of cold pool frequency in terms of ASCAT while for ICON, the model did not produce strong, deep convection over the Atlantic (Stevens et al., 2019), thus leading to smaller number of TFs in the basin. In the case of IO, RapidScat GFs lead in relative frequency, very closely trailed by ASCAT GFs and then ICON TFs. Convection over the Indian ocean is dominated by the frequent passage of monsoonal intraseasonal oscillations, MJO, and boreal summer intraseasonal oscillations (BSISO). Both ICON and ASCAT depict a similar number of cold pools, thus suggesting that ICON was able to simulate mesoscale convection closer to long-term mean. The slightly higher number of cold pools depicted by RapidScat over Indo-Pacific warm pool may have been due to relatively stronger convective activity as observed in Dong and McPhaden (2018). Overall, the comparison suggests that ICON did not significantly deviate from either ASCAT or RapidScat in terms of producing total number of cold pools in each basin except over the Atlantic, thus suggesting the applicability of this simulation in understanding the relationship between mesoscale cold pools and their respective environment in the coming sections.

In addition to the cold pool frequency, comparing the size of cold pools and precipitation between the simulation and satellite observations will help understand how well the model can produce cold pool properties. Figure 6 shows the percentile range from 5^{th} to 99^{th} percentiles of size and precipitation for ASCAT, RapidScat, and ICON where the increment is 5% between 5 and 95% and 4% between 95-99%. Comparing Fig. 6a-b, it can be seen that ICON TF sizes compare well with both RapidScat (Fig. 6a) and ASCAT (Fig. 6b) with IO exhibiting the largest TFs and GFs for ICON and RapidScat, followed by tropical average, WPAC, AO, and IO. ASCAT-ICON size distribution is similar to RapidScat-ICON except that IO exhibits a smaller cold pool size for ASCAT. This size comparison suggests that although ICON-simulated cold pools are smaller than satellite-derived cold pools, the model reproduced inter-basin variability well. In the case of precipitation, Fig. 6c-d suggests that the percentile distribution is more linear than for size, especially in tropical average and EPAC, thus suggesting a quasi-monotonic relationship between TF and GF precipitation. Note that ICON TF precipitation is greater in range than both RapidScat and ASCAT GF precipitation. WPAC and IO have the highest precipitation in both Fig. 6c-d, thus implying that ICON-simulated cold pools within Indo-Pacific warm pool convection have similar distribution as scatterometer-identified cold pools and their parent convection. ICON relatively overestimates precipitation as compared to RapidScat over the AO (Fig. 6c) and ASCAT over the EPAC (Fig. 6d). This could be due to the frequent tropical easterly wave-attributed convection produced in these basins, which is further discussed below. Note that TFs are smaller but rainier than the GFs. The hypothesis behind smaller TFs is that as cold pools move over the ocean, the temperature deficit may recover while the wind gradient signal in GFs can persist longer, thus resulting in a much larger and longer-lasting outflow boundary. Previous studies have shown that cold pools can last longer than their parent convection (Zuidema et al., 2017; Grant et al., 2018; Vogel et al., 2021) and thus it is possible that scatterometers are able to observe the GFs which have moved away from their respective parent systems (Garg et al., 2020). Also, the difference in spatial resolution of ICON and RapidScat/ASCAT would be another reason that ICON represents smaller cold pools than either scatterometer's GFs. In the case of daily precipitation, Stevens et al. (2020) showed that ICON, even at 2.5 km, overestimated precipitation as SRM suffers in distinguishing shallow and deep convective regimes over open oceans as compared to observations. However, ICON TFs are able to represent the inter-basin spread and variability in frequency, size, and precipitation well.

Now that global TF properties are identified, the diurnal cycle of TF number density and other attributed properties will be analyzed in the following section.

4 Diurnal Cycle of ICON-Simulated Cold Pools

The diurnal cycle of convection and precipitation is one of the most important modes of variability across the global land and oceanic regions (e.g., Chapman, 1951; Haurwitz, 1964; Brier, 1965). Several observational studies from a range of remote sensing and *in situ* observations have provided evidence of an early morning/late night peak and an afternoon peak in precipitation over global tropical oceans (Gray & Jacobson, 1977; Reed & Jaffe, 1981; Albright et al., 1985; Augustine, 1984; Nesbitt & Zipser, 2003; Kikuchi & Wang, 2008). Although the models have improved the representation of the diurnal variation of precipitation due to improvement in the parameterization of convective processes, there are still biases in model-simulated diurnal peaks in terms of intensity and timing (e.g., Giles et al., 2020; Wei et al., 2020). This section depicts the diurnal cycle of ICON-simulated cold pool properties and compares it with the climatological diurnal cycle of RapidScat-observed GF number density (Garg et al., 2021) and IMERG-observed precipitation. All diurnal analysis in this section is performed within the four boxes of the western Pacific (WPAC), eastern Pacific (EPAC), Atlantic (AO), and Indian Ocean (IO) as depicted in Fig. 1 (Garg et al., 2021).

The diurnal cycle of the number of cold pools from ICON (solid blue; 40-day), RapidScat (solid red; 2014-2016), and IMERG precipitation (dotted purple; 2000-2020) is shown in Fig. 7 for all four boxes and the global tropical average. First, comparing the ICON TF with RapidScat GF diurnal cycle, it can be seen that TFs show an early morning/late night peak (0000 - 0600 LT) similar to RapidScat but with different timing and intensity for all basins and the tropical average. Second, ICON does not exhibit the afternoon peak in cold pool number as prominently as RapidScat for all basins except AO, where ICON shows a daytime peak at noon compared to RapidScat peak at 1600 LT. Third, IMERG precipitation does not show an afternoon maxima in precipitation corresponding to RapidScat GF number density. From the conceptual framework created in Garg et al. (2021) using RapidScat data, the early morning/late night peak is related to deep, organized moist convection (e.g., MCSs) while the afternoon peak pertains to the shallow cumulus congestus type of convection. Applying this conceptual framework to ICON-simulated TF number density and IMERG precipitation here, this analysis suggests that ICON and IMERG are missing precipitation from the cumulus congestus type of convection in general but can capture the TFs associated with mesoscale organized convection (e.g., MCSs) during night time/early morning. This result is consistent with (Stevens et al., 2019) showing a lack of shallow convection in the simulation.

In addition to cold pool number, looking at the diurnal cycle of cold pool-attributed ambient environmental properties shown in Fig. 1-3 will help understand how environmental properties change during the day associated with cold pool activity. Figure 8 shows the interquartile range (IQR) of different cold pool-associated parameters. Overall, most of the parameters show a weak or moderate diurnal variation except TF size (Fig. 8a), precipitation (Fig. 8f), and CAPE (Fig. 8g). TF size and CAPE have maxima in the afternoon between 1200-1800 LT, while TF precipitation has an early morning peak between 0000 - 0600 LT (similar to TF frequency in Fig. 7). Comparing TF number density (Fig. 7), size (Fig. 8a), and TF precipitation (Fig. 8f) with RapidScat-observed GF number, size, and precipitation in Fig. 10 of Garg et al. (2021), both have similar temporal co-variation as precipitation (late night) peaks first followed by cold pool frequency (early morning) and size (afternoon). Also, TF size peaks in the late morning/afternoon in both RapidScat (Garg et al., 2021) and ICON-observed cold pools, suggesting the role of upscale growth during the overnight and early morning hours resulting in the formation of new cold pools which can intersect and merge to form bigger cold pools later in the day. Previous studies have observed a maximum in CAPE during the afternoon-to-evening hours (e.g., Dai et al., 1999; Bechtold et al., 2004) as the sea surface warms in the presence of incoming solar radiation, resulting in stronger surface fluxes and CAPE in the environment. The similarity between TF size and CAPE diurnal cycle suggests that CAPE, in particular, can prove to be an important parameter in predicting TF size in convective parameterizations.

To summarize, the diurnal cycle of ICON-simulated cold pool number density, size, and associated properties do show similar early morning peak timing with respect to RapidScat-observed GF properties but miss the afternoon peak in TF frequency. This can be because ICON at 2.5 km resolution might misrepresent shallow convective regions as it has been observed in multiple studies in the past that CRMs, even at kilometer resolution, can improperly represent narrow updrafts associated with shallow convective clouds (Bryan & Morrison, 2012; Varble et al., 2014). Therefore, ICON can identify mesoscale downdrafts formed near MCS-type systems during the late night/early morning hours but may miss the afternoon signal in cold pool activity formed due to congestus-type systems. However, this should not affect the analysis of nocturnal deep convective cold pools. The next section presents the relationship between environmental properties and cold pool number, size, and intensity for mesoscale cold pools using a ML method.

5 Relationship between Environment and Cold Pool Properties

Statistical methods such as simple and multiple linear regression have been used to identify and analyze the relationships between environmental conditions and convection-precipitation (Jung et al., 2010; Goyal et al., 2014). However, due to the nonlinearity in environmental variables, linear regression can underfit, resulting in difficulty deriving relationships between the convection and its environment. Random forest (RF) regression and classification are based on decision trees (Breiman, 2001; Hastie et al., 2009) which produces numerous independent trees to obtain a final decision via two randomization approaches. One is through the selection of training samples and the other is by selecting important variables at each node of a tree. This randomization reduces typical drawbacks of decision trees such as overfitting problem and sensitivity of the output to the training sample configuration (Breiman, 2001). RF also provides the option of using out-of-bag (OOB) samples from random selection to provide internal cross-validation and relative importance of a variable when samples are drawn from the OOB (e.g., Stumpf & Kerle, 2011; Long et al., 2013; Kim et al., 2014; Maxwell et al., 2014). The final prediction from the RF approach is an average over all the trees used in the algorithm. O'Gorman and Dwyer (2018) used a random forest algorithm to replace convective parameterization in a GCM and observed that the RF approach was able to preserve physical constraints such as energy conservation, thus leading to accurate and stable simulations of climate in a GCM. In this study, a random forest regression approach is implemented to identify which environmental properties are most related to cold pool density, size, and intensity across the deep global tropics. All the analysis is carried out for deep tropics (-15^o to 15^o in latitude) to identify the most relevant features in the regions of moist vigorous convection. Also, TFs with 2m air temperature less than 293 K, precipitation less than 1 mm hr^{-1} and size less than 1000 km^2 (all three conditions should be satisfied together) are removed from the analysis as it is hypothesized that these TFs are not related to the deep moist tropical convection after locating most of these systems in the Austral winter hemisphere (not shown). In this way, approximately 15 million TFs are used in this analysis. All the features and predictands were standardized using *z-scores* in order to normalize the scale of the values. After applying the mentioned criterion, the training and validation sample is split 70-30 from the entire TF dataset used in this study.

5.1 Random Forest Algorithm Structure

Hyperparameters in an ML algorithm are configurations that are external to the ML model being used and cannot be derived from the data itself. In the case of RF, some hyperparameters are important to consider to improve the accuracy of any model. For example, the higher the number of trees, the better the results in terms of performance and precision, but for some predictive problems, adding trees diminishes the improvement in model accuracy. Other hyperparameters such as node size and sample size control the randomness of the RF and thus need to be properly evaluated before using the model for any classification or regression problem (Probst et al., 2019). Hyperparameter tuning is a method to obtain the combination of these configurations to obtain the best possible accuracy in the model. This study uses the python-based *scikit-learn* scikit-learn random forest regression tool to implement the RF algorithm. To train the RF algorithm, all the environmental features from ICON (TCWV, CAPE, surface - 600 hPa and surface - 300 hPa wind shear, RH, wind gradient, precipitation, SHF, and LHF) in a 0.5^o grid box (Fig. 1-3) are used at this step to predict TF number density, size, and intensity (T_v anomaly). Scikit-learn's *RandomizedSearchCV* (Pedregosa et al., 2011) method is first implemented on a permutation of a number of estimators, maximum features, maximum depth of the tree, minimum samples used to split, minimum samples on each leaf, and bootstrap configuration by running 100 iterations on a 3-fold cross-validation matrix. What this step essentially does is that it provides a combination of hyperparameter configuration, which has led to the most accurate prediction, but it is still randomized. To narrow down the hyperparameter space further, scikit-learn's *GridSearchCV* (Pedregosa et al., 2011) is applied to a narrow set of hyperparameters obtained from a random search in the previous step. *GridSearchCV* searches through every combination of the configuration provided to refine the results further and provide us with the most accurate result with low overfitting. This exercise provided values of hyperparameters which resulted in high R^2 (correlation) and low root mean square error (RMSE) between predicted and true values of TF number density, size, and intensity. Table 3 shows the values of hyperparameters used for the RF algorithm in this study.

5.2 Training the RF Regression Algorithm

Unlike the multiple linear regression models, RF does not provide coefficients corresponding to each predictor used in the training algorithm since the nature of the random forest algorithm inherently leads to the destruction of any simple mathematical formulation as RF works by building decision trees and then aggregating them (Breiman, 2001). Therefore, to look at the working mechanism of the RF model, *feature importance* needs to be analyzed to understand which features affect predictand the most. Figure 9 provides the importance of all the features used in this study relevant for TF number density (Fig. 9a), size (Fig. 9b), and intensity (Fig. 9c). For frequency of TFs in Fig. 9a, TCWV, precipitation, CAPE, surface - 600 hPa, and surface - 300 hPa vertical wind shear are the five most important features whose relative importance sum up to be approximately 80%. Similarly, for TF size (Fig. 9b), TCWV, precipitation, CAPE, RH, and surface - 600 hPa wind shear contribute the most (approx 80%). Going back to Fig. 6 in Garg et al. (2021), where RapidScat-observed GF number density shows a strong correlation with TCWV and relating it to Fig. 9a where TCWV exhibits the strongest control on TF number density, it can be said that the total moisture present in the vertical column of the parent convective system is vital in characterizing cold pool number at the surface, as the occurrence of deep moist convection is shown to depend on TCWV (Schiro et al., 2016; Schiro & Neelin, 2019) and should have stronger downdraft mass flux. Similarly, TCWV shows the highest relative importance for TF size, thus suggesting the role of column moisture in providing positive feedback to cold pool production, which can merge to form larger cold pools at the surface. Precipitation has the second most important role for TF number density and size, as heavier precipitation should lead to more downdrafts at the surface, thus resulting in numerous and larger cold pools. CAPE is the third most relevant feature for both TF frequency and size, and as we saw in the global TF climatologies in Fig. 1-2, a higher CAPE coincides well with moister, numerous, and larger TFs. For the number of TFs, the next two important features are mid-tropospheric and upper-tropospheric wind shear. The Wind shear climatology in Fig. 3c-d shows that low-to-moderate wind shear is prominently observed over most deep, moist tropical convective regimes. This finding suggests that most of the tropical oceanic cold pools are formed in low-to-moderate shear environments. TF size has a similar weight to surface - 600 hPa wind shear as the number of TFs. Among the five most important features, the only difference between TF frequency and size is RH. It shows a higher weight than wind shear in controlling TF size, thus suggesting the importance of net available moisture in modulating cold pool size.

In the case of TF intensity in Fig. 9c, CAPE has the strongest control followed by RH, wind gradient, LHF, and TCWV (total approx 82%). Comparing T_v anomaly in Fig. 2a with wind gradient (Fig. 2b), RH (Fig. 2c), LHF (Fig. 3b), and TCWV (Fig. 1d), coldest cold pools are observed in the areas of high CAPE, moisture, LHF, and wind gradient. Thus it makes sense that the RF algorithm has identified these features as the most important in controlling the intensity of cold pools at the surface. Physically, it suggests that when the environment has more moisture available, it should lead to higher CAPE as it is a function of temperature and moisture. These conditions should lead to the formation of deep convection with heavier precipitation, thus leading to colder T_v anomalies at the surface with strong wind gradients at their boundaries. Since LHF is essentially a function of moisture and wind speeds, it should be directly related to the cold pool-attributed changes. In this way, RF captured physically constrained environmental features important for cold pool activity over the global tropical oceans.

5.3 Validation of Trained RF Algorithm

After the RF regression learned the relationship between the environmental parameters and cold pool properties, validation is carried out on an independent dataset (70-30 split) to obtain the R^2 score and RMSE between true and predicted values, where predicted values are RF regression generated values of cold pool number density (or frequency), size, and T_v anomaly using the features in Fig. 9. Figure 10 shows the scatter plot of true versus predicted values of cold pool number density, size, and T_v anomaly using all the features in Fig. 9 and using the 5 most important features as well. Comparing the RMSE and R^2 score between the scatter plots using all the features and the 5 most important features, it can be seen that Fig. 10d-f exhibits minor underperformance as compared to Fig. 10a-c. This suggests that although the remaining environmental features are important to produce accurate results, the 5 most important predictors are sufficient enough to predict the cold pool frequency, size, and T_v anomaly with high precision (low RMSE and high R^2). Note that the RF algorithm performs the best in the case of TF frequency followed by intensity and size in both cases (Fig. 10a-c and Fig. 10d-f). This order suggests that RF regression learned the relationship between the environmental conditions and TF frequency and intensity better than the size. Garg et al. (2021) showed that RapidScat-observed GF size has a highly nonlinear relationship with GF number frequency, minimum brightness temperature (T_B), and TCWV, as intersecting cold pools could play a significant role in defining the cold pool size as observed in Feng et al. (2015). On the other hand, RapidScat-observed GF frequency had a strong quasi-linear relationship with both minimum T_B and TCWV, suggesting a strong control of convection type and moisture present in the environment. Applying similar reasoning as above, TF frequency and intensity portray a strong relationship with the environment compared to size, which could have resulted in better performance of RF for these two cold pool properties.

6 Exploring Regional Relationships between Features and Cold Pool Properties

6.1 Correlation Matrix of Environment and Cold Pool Properties

Understanding regional differences in the relationship between environmental properties and cold pool characteristics are important to improve the physical parameterizations in the current weather and climate models. Even though the RF regression algorithm could interpret the physical relationships reasonably well across the global tropics, regional differences in such statistics need to be analyzed to identify differences in convective processes within different ocean basins. Therefore, a Pearson's correlation coefficient matrix at 95% significance level is presented in Fig. 11 for global tropical regions, deep tropics (as defined in the previous section), WPAC, EPAC, IO, and AO. Note that the white boxes in the lower triangular matrix signify no statistically significant relationship between the variables. Table 4 shows the sample size of TF grid points used to calculate these correlation values in each panel.

Comparing the global tropics (Fig. 11a) with the deep tropics as defined in the previous section (Fig. 11b), all features behave similarly with minor changes in correlation values between the two. However, over the Pacific Ocean, comparing WPAC (Fig. 11c) and EPAC (Fig. 11d), there are major differences in the correlation values. For instance, the number of TFs and TF size over WPAC have negative correlation values with TF precipitation compared to EPAC and all other regions shown in Fig. 11. Also, TF precipitation and TCWV negatively correlate with CAPE for WPAC, while both these variables exhibit a strong positive relationship for all other regions. The number of TFs and TF size also negatively correlate with RH over WPAC, while they positively correlate with RH over other regions. In the scatter matrix for WPAC (not shown), TF frequency and size depict a narrow distribution with TF precipitation and RH, and thus although they are not necessarily negatively correlated, a weak negative correlation is present in Fig. 11c. Similarly, TCWV and TF precipitation vary within a narrow distribution with CAPE, thus depicting a negative correlation in Fig. 11c. EPAC (Fig. 11d), on the other hand, shows similar correlation values as deep (Fig. 11a) and global tropics (Fig. 11b).

AO in Fig. 11e depicts a weakly positive correlation between T_v anomaly and TF density, size, precipitation, and TCWV, which is opposite to other basins and global tropics. AO also depicts a weak negative correlation between wind gradient and TF number density, size, precipitation, and TCWV. The scatter matrix for AO (not shown) suggests that T_v anomaly has a very narrow distribution around -2 K for these variables, which is not necessarily negatively correlated with the variables mentioned above. AO (Fig. 11e) also depicts a weak negative correlation between surface fluxes and TF size, precipitation, TCWV, and CAPE. The Scatter matrix further signifies that the relationship between SHF, LHF, and TF properties is nonlinear in nature, with most of the values concentrated in a narrow range, thus resulting in a weak negative correlation between these parameters.

IO in Fig. 11f shows relatively similar values as global and deep tropics (Fig. 11a-b) with an exception in surface - 300 hPa wind shear showing a positive correlation with TF number density, size, precipitation, TCWV, and CAPE compared to the negative relationship in the global and deep tropics. Comparing this result with global climatology of surface - 300 hPa wind shear in Fig. 3d, it can be seen that the Indian Ocean has relatively high upper tropospheric wind shear related to the TEJ, which is quite prominent during the South Asian monsoon regime. Therefore, it makes physical sense that most of the TFs are related to the relatively high upper tropospheric wind shear over IO. Hypothesized physical mechanisms explaining these differences are presented in the following section.

6.2 Physical Mechanisms associated with Regional Difference in TF Properties

Regional differences in the relationships between TF and environmental properties were explored in the previous section using statistical methods. However, it is important to physically interpret the reasoning behind these regional differences to guide the physical parameterizations in the climate models. To understand the controls on cold pool properties over different basins, a Hovmöller diagram between longitude and time (binned by day) of TF number frequency anomaly (green markers) and 860 hPa meridional wind anomaly is shown in Fig. 12 for WPAC (Fig. 12a), EPAC (Fig. 12b), AO (Fig. 12c), and IO (Fig. 12d). Figure 12 also depicts the time series of TF frequency for each basin on top of each Hovmöller diagram to identify the periodicity of TF occurrence. The anomalies are calculated by subtracting the mean along each longitude bin for all the time steps. Note that a weak MJO according to Realtime Multi-variate MJO (RMM) index was over the western Pacific and the Indian Ocean during the 40-day ICON simulation. Hovmöller diagram here can provide information about the wave activity governing the TFs in the model simulation.

Comparing the four basins in Fig. 12, WPAC and IO show similar characteristics in Hovmöller diagrams and time series of TF distribution across longitude and time. Chen and Houze (1997) observed that the Indo-Pacific warm pool is the region of intense deep convection with longer lifetimes, and local-mesoscale forcing is dominant for convective initiation, sustenance, and longevity. Relating the observations of Chen and Houze (1997) with Fig. 12, it can be said that deep moist convection over the IO and WPAC is ubiquitous, leading to numerous cold pools at the surface, which would provide positive feedback to secondary convection. The time series of the number of TFs in Fig. 12 further shows that the number of TFs over WPAC (Fig. 12a) is relatively consistent over time with a minor temporal variation. IO (Fig. 12d) does exhibit an intra-seasonal variability in the TF frequency, which could be related to the active and break period of the boreal summer monsoon (Ramamurthy, 1969; Krishnan et al., 2000; Lawrence & Webster, 2002; Webster & Hoyos, 2004; Pattanaik et al., 2020). IO also depicts the BSISO in meridional wind anomaly variability consistent with the patterns associated with the BSISO in Fig. 12d represented by blue patch around 08 Aug and then dark red patch around 02 Sep 2016. Even though the intraseasonal oscillation is visible over the IO, mesoscale controls due to longer-lived MCS type of systems play an important role in producing convective cold pools over the Indian Ocean as shown in Huang et al. (2018) and Roca and Fiolleau (2020). Fig. 12 therefore suggests that cold pools over the Indo-Pacific basin have a robust control through local mesoscale circulations with relatively less synoptic forcing.

In the case of AO in Fig. 12c, the number of TFs exhibits a signature of tropical easterly waves (frequency of 3-5 days), which is one of the most common synoptic modes of variability over the Atlantic Ocean (e.g., Carlson, 1969; Reed et al., 1977; Avila & Pasch, 1995; Kiladis et al., 2006; Mekonnen & Rossow, 2011). Time series in Fig. 12c further provides evidence of easterly waves-associated TF activity over the AO. EPAC (Fig. 12b) shows a similar easterly wave-related TF signature with the time series showing 8-10 days variability. Previous studies have attributed convection over the eastern tropical Pacific to easterly wave structures (Riehl, 1954; Yanai et al., 1968; Burpee, 1972; Molinari et al., 1997; D. D. Raymond et al., 1998; Zehnder et al., 1999; Molinari et al., 2000; Gu & Zhang, 2002; Serra & Houze Jr, 2002) during the boreal summer. Thus, it can be said that EPAC and AO TFs are primarily controlled through the synoptic easterly wave circulation compared to WPAC and IO, where mesoscale circulations play a crucial role in the convective system initiation and sustenance during the model simulation period.

7 Summary and Conclusions

Although cold pools and their associated characteristics have been analyzed in the past using a range of observational and modeling frameworks, a holistic global tropical oceanic cold pool analysis comparing global SRM and observations was yet to be carried out. With this motivation, this study used T_v anomaly to identify TFs in the ICON model for a 40-day simulation and identified approximately 16.5 million cold pools. The global simulated climatology of TF-attributed environmental properties (e.g., number density, size, precipitation, CAPE, wind shear, TCWV, etc.) was created and analyzed to examine the relationship between tropical oceanic cold pool activity and their environments. TF frequency matched well with ASCAT- (Garg et al., 2020) and RapidScat-observed (Garg et al., 2021) cold pool frequency climatologies. A comparative analysis between 12-year ASCAT and 2-year RapidScat GF climatology with ICON TF climatology shows that Indo-Pacific and western Atlantic identified more TFs than GFs due to enhanced mesoscale activity in the model during the boreal summer. Comparison of size and precipitation percentiles between ICON, ASCAT, and RapidScat further shows that model-simulated cold pools had similar inter-basin variability as the observed climatology but with different ranges of size and precipitation.

The diurnal cycle of ICON-simulated TF frequency was compared against RapidScat-observed GF frequency. It was found that the ICON was able to simulate the early morning/late night peak in TFs but missed the afternoon maxima in cold pool activity as observed in the RapidScat-associated GF climatology in Garg et al. (2021). These results are summarized through an illustration in Fig. 13; the top panel shows the characteristics of ICON-cold pool and IMERG-precipitation diurnal cycle, while the bottom panel depicts the RapidScat-observed GF diurnal cycle. The observed cold pool diurnal cycle in RapidScat (Fig. 13b) suggests that the observed shallow congestus convection produces cold pools of mesoscale dimension. New convection triggering and surface-atmosphere exchanges in the vicinity of cold pools may provide the necessary ingredients for upscale growth into the evening, growing into mesoscale-organized deep convection overnight (e.g., MCSs). These MCSs further intensify cold pool number and cold pool areal coverage, likely enhancing surface-atmosphere moisture and energy exchanges. On the other hand, ICON (Fig. 13a) depicts MCS-type convection overnight similar to RapidScat (Fig. 13b) but it lacks congestus-type convection, which has implications for how the model represents both cloud-radiative interactions as well as surface-atmosphere exchanges in the afternoon prior to the evening oceanic upscale growth period. This study also finds that despite observations of congestus clouds in infrared and GF observations, IMERG precipitation retrievals also lack an afternoon precipitation peak associated with mid-level convection.

Systematic biases exist in the representation of cloud types critical for vertical transports and latent heating, radiative forcing, and the organization of convection within the tropical diurnal radiative cycle. Therefore, it is an important topic of future work which should aim to better gauge how radiation, boundary layer, and microphysics schemes in storm resolving models such as the ICON model are working together to represent key aspects of the diurnal cycle, including the shallow-to-deep convective transition and convective organization, and their effects on the tropical water cycle and energy budgets. It is expected that global SRMs may lack the spatial resolution to produce congestus convection at kilometer scales (e.g., Varble et al., 2014), however significant and insidious biases in weather and climate models resulting from misrepresenting the key processes involved in congestus type convection should be further investigated.

Once it was established that ICON could represent nocturnal mesoscale convective cold pools well, a random forest regression algorithm was applied to environmental features (CAPE, mid-and upper tropospheric wind shear, TCWV, RH, wind gradient, precipitation, SHF, and LHF) to identify their relationship with TF frequency, size, and intensity. RF was trained on 0.5^{o} gridded composited TF dataset integrated over the entire simulation and was validated against an independent subset of the data. RF performed well in predicting TF frequency, size, and intensity using the five most important features in each case and thus provided valuable information about environmental controls on TF properties over the deep tropical oceans with TCWV, precipitation, wind shear, and CAPE proving to be the most important environmental properties relevant for cold pool activity over the deep tropics. Pearson's correlation coefficients were calculated at a 95% significance level between all environmental features mentioned above, and TF properties for WPAC, EPAC, IO, and AO basins were compared with the global and deep tropics. Regional differences in TF relationships were analyzed, and physical explanations were provided using the Hovmöller diagram (time versus longitude) of the number of TFs overlayed with 860 hPa meridional wind anomalies. In addition, the time series of TF frequency was calculated for each basin and compared with each other to identify the periodicity of TF frequency during the entire simulation. It was observed that WPAC and IO have a strong control through mesoscale circulations formed locally in addition to synoptic forcing from boreal summer monsoon and BSISO. On the other hand, EPAC and AO TF activity is primarily controlled through synoptic forcing from tropical easterly wave structures. These findings match well with the previous studies carried out over these basins using a range of observational and model frameworks. However, a more robust analysis (e.g., analyzing the relationship of downdraft mass flux with the convective environment and the role of microphysics in controlling cold pool activity at the surface) is required to understand further the relationship between cold pools and their parent environment to improve the representation of convection in the models. Overall, this study provided a holistic comparison between observations and GSRMs, which are hypothesized to be a bridge between observations and climate models, to identify the biases in GSRMs to improve the further representation of convection and precipitation in global weather and climate models.

Acknowledgments

The authors would like to thank the NASA Ocean Vector Wind Science Team for providing funding through grant NNH17ZDA001N-OVWST 17-OVWST-17-0019. The authors also would like to thank the NASA PODAAC team for providing RapidScat data. The authors also want to thank NASA GES DISC for providing IMERG data used in this study. DYAMOND data management was provided by the German Climate Computing Center (DKRZ) and supported through the projects ESiWACE and ESiWACE2. The projects ESiWACE and ESiWACE2 have received funding from the European Union's

Horizon 2020 research and innovation programme under grant agreements No 675191 and 823988. This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project IDs bk1040 and bb1153 on Mistral supercomputing cluster. The data used in this study is provided in a public domain at https://publish.illinois.edu/scat-coldpools/data-products/ and is available for public use.

References

Albright, M. D., Recker, E. E., Reed, R. J., & Dang, R. (1985, 10). The Diurnal Variation of Deep Convection and Inferred Precipitation in the Central Tropical Pacific During January–February 1979. *Monthly Weather Review*, *113*(10), 1663-1680. doi: 10.1175/1520-0493(1985)113⟨1663:TDVODC⟩2.0.CO;2

Augustine, J. A. (1984). The diurnal variation of large-scale inferred rainfall over the tropical pacific ocean during august 1979. *Monthly weather review*, *112*(9), 1745–1751.

Avila, L. A., & Pasch, R. J. (1995). Atlantic tropical systems of 1993. *Monthly weather review*, *123*(3), 887–896.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., & Reinhardt, T. (2011). Operational convective-scale numerical weather prediction with the cosmo model: Description and sensitivities. *Monthly Weather Review*, *139*(12), 3887–3905.

Bechtold, P., Chaboureau, J.-P., Beljaars, A., Betts, A., Köhler, M., Miller, M., & Redelsperger, J.-L. (2004). The simulation of the diurnal cycle of convective precipitation over land in a global model. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *130*(604), 3119–3137.

Bhat, G., Srinivasan, J., & Gadgil, S. (1996). Tropical deep convection, convective available potential energy and sea surface temperature. *Journal of the Meteorological Society of Japan. Ser. II*, *74*(2), 155–166.

Borque, P., Nesbitt, S. W., Trapp, R. J., Lasher-Trapp, S., & Oue, M. (2020). Observational study of the thermodynamics and morphological characteristics of a midlatitude continental cold pool event. *Monthly Weather Review*, *148*(2), 719–737.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, *45*(12), 6289-6298. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL078510` doi: https://doi.org/10.1029/2018GL078510

Bretherton, C. S., Peters, M. E., & Back, L. E. (2004). Relationships between water vapor path and precipitation over the tropical oceans. *Journal of climate*, *17*(7), 1517–1528.

Brier, G. W. (1965). Diurnal and semidiurnal atmospheric tides in relation to precipitation variations. *Monthly Weather Review*, *93*(2), 93–100.

Brooks, H. E., Doswell III, C. A., & Cooper, J. (1994). On the environments of tornadic and nontornadic mesocyclones. *Weather and forecasting*, *9*(4), 606–618.

Bryan, G. H., & Morrison, H. (2012). Sensitivity of a simulated squall line to horizontal resolution and parameterization of microphysics. *Monthly Weather Review*, *140*(1), 202–225.

Burpee, R. W. (1972). The origin and structure of easterly waves in the lower troposphere of north africa. *Journal of the Atmospheric Sciences*, *29*(1), 77–90.

Carlson, T. N. (1969). Some remarks on african disturbances and their progress over the tropical atlantic. *Monthly Weather Review*, *97*(10), 716–726.

Chandra, A. S., Zuidema, P., Krueger, S., Kochanski, A., de Szoeke, S. P., & Zhang, J. (2018). Moisture distributions in tropical cold pools from equatorial indian ocean observations and cloud-resolving simulations. *Journal of Geophysical Research: Atmospheres*, *123*(20), 11–445.

Chapman, S. (1951). Atmospheric tides and oscillations. In *Compendium of meteorology* (pp. 510–530). Springer.

Charba, J. (1974). Application of gravity current model to analysis of squall-line gust front. *Monthly Weather Review*, *102*(2), 140–156.

Chen, S. S., & Houze, R. (1997). Interannual variability of deep convection over the tropical warm pool. *Journal of Geophysical Research: Atmospheres*, *102*(D22), 25783–25795.

Cheng, W.-Y., Kim, D., Rowe, A., Moon, Y., & Wang, S. (2020). Mechanisms of convective clustering during a 2-day rain event in amie/dynamo. *Journal of Advances in Modeling Earth Systems*, *12*(3), e2019MS001907. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001907` (e2019MS001907 2019MS001907) doi: https://doi.org/10.1029/2019MS001907

Cortinas, J. V., & Stensrud, D. J. (1995). The importance of understanding mesoscale model parameterization schemes for weather forecasting. *Weather and forecasting*, *10*(4), 716–740.

Cotton, W. R., Alexander, G. D., Hertenstein, R., Walko, R. L., McAnelly, R. L., & Nicholls, M. (1995). Cloud venting—a review and some new global annual estimates. *Earth-Science Reviews*, *39*(3-4), 169–206.

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

de Szoeke, S. P., Skyllingstad, E. D., Zuidema, P., & Chandra, A. S. (2017). Cold pools and their influence on the tropical marine boundary layer. *Journal of the Atmospheric Sciences*, *74*(4), 1149–1168.

Diffenbaugh, N. S., Scherer, M., & Trapp, R. J. (2013). Robust increases in severe thunderstorm environments in response to greenhouse forcing. *Proceedings of the National Academy of Sciences*, *110*(41), 16361–16366.

Dong, L., & McPhaden, M. J. (2018). Unusually warm indian ocean sea surface temperatures help to arrest development of el niño in 2014. *Scientific reports*, *8*(1), 1–10.

Feng, Z., Hagos, S., Rowe, A. K., Burleyson, C. D., Martini, M. N., & de Szoeke, S. P. (2015). Mechanisms of convective cloud organization by cold pools over tropical warm ocean during the amie/dynamo field campaign. *Journal of Advances in Modeling Earth Systems*, *7*(2), 357-381. doi: 10.1002/2014MS000384

Feng, Z., Leung, L. R., Liu, N., Wang, J., Houze Jr, R. A., Li, J., … Guo, J. (2021). A global high-resolution mesoscale convective system database using satellite-derived cloud tops, surface precipitation, and tracking. *Journal of Geophysical Research: Atmospheres*, e2020JD034202.

Fuhrer, O., Chadha, T., Hoefler, T., Kwasniewski, G., Lapillonne, X., Leutwyler, D., … others (2018). Near-global climate simulation at 1 km resolution: establishing a performance baseline on 4888 gpus with cosmo 5.0. *Geoscientific Model Development*, *11*(4), 1665–1681.

Gallego, D., Ribera, P., Garcia-Herrera, R., Hernandez, E., & Gimeno, L. (2005). A new look for the southern hemisphere jet stream. *Climate Dynamics*, *24*(6), 607–621.

Garg, P., Nesbitt, S. W., Lang, T. J., & Priftis, G. (2021). Diurnal cycle of tropical oceanic mesoscale cold pools. *Journal of Climate*, 1 - 60. Retrieved from `https://journals.ametsoc.org/view/journals/clim/aop/JCLI-D-20-0909.1/JCLI-D-20-0909.1.xml` doi: 10.1175/JCLI-D-20-0909.1

Garg, P., Nesbitt, S. W., Lang, T. J., Priftis, G., Chronis, T., Thayer, J. D., & Hence, D. A. (2020). Identifying and characterizing tropical oceanic mesoscale cold pools using spaceborne scatterometer winds. *Journal of Geophysical Research: Atmospheres*, *125*(5), e2019JD031812. doi: 10.1029/2019JD031812

Gassmann, A., & Herzog, H.-J. (2008). Towards a consistent numerical compressible non-hydrostatic model using generalized hamiltonian tools. *Quarterly Journal of the Royal Meteorological Society*, *134*(635), 1597–1613.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, *45*(11), 5742-5751. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL078202` doi: https://doi

.org/10.1029/2018GL078202

Giles, J. A., Ruscica, R. C., & Menéndez, C. G. (2020). The diurnal cycle of precipitation over south america represented by five gridded datasets. *International Journal of Climatology*, *40*(2), 668–686.

Goyal, M. K., Bharti, B., Quilty, J., Adamowski, J., & Pandey, A. (2014). Modeling of daily pan evaporation in sub tropical climates using ann, ls-svr, fuzzy logic, and anfis. *Expert systems with applications*, *41*(11), 5267–5276.

Grant, L. D., Lane, T. P., & van den Heever, S. C. (2018). The role of cold pools in tropical oceanic convective systems. *Journal of the Atmospheric Sciences*, *75*(8), 2615 - 2634. Retrieved from `https://journals.ametsoc.org/view/journals/atsc/75/8/jas-d-17-0352.1.xml` doi: 10.1175/JAS-D-17-0352.1

Gray, W. M., & Jacobson, R. W. (1977, 09). Diurnal Variation of Deep Cumulus Convection. *Monthly Weather Review*, *105*(9), 1171-1188. doi: 10.1175/1520-0493(1977)105⟨1171:DVODCC⟩2.0.CO;2

Gu, G., & Zhang, C. (2002). Cloud components of the intertropical convergence zone. *Journal of Geophysical Research: Atmospheres*, *107*(D21), ACL–4.

Guichard, F., & Couvreux, F. (2017). A short review of numerical cloud-resolving models. *Tellus A: Dynamic Meteorology and Oceanography*, *69*(1), 1373578.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Random forests. In *The elements of statistical learning* (pp. 587–604). Springer.

Haurwitz, B. (1964). Comments on wave forms in noctilucent clouds, geophys. *Institute Sci. Rep. No. UAGR-160*.

Hohenegger, C., Kornblueh, L., Klocke, D., Becker, T., Cioni, G., Engels, J. F., ... Stevens, B. (2020). Climate statistics in global simulations of the atmosphere, from 80 to 2.5 km grid spacing. *Journal of the Meteorological Society of Japan. Ser. II*, *98*(1), 73-91. doi: 10.2151/jmsj.2020-005

Holloway, C. E., & Neelin, J. D. (2009, 06). Moisture Vertical Structure, Column Water Vapor, and Tropical Deep Convection. *Journal of the Atmospheric Sciences*, *66*(6), 1665-1683. doi: 10.1175/2008JAS2806.1

Houze, Rasmussen, K. L., Zuluaga, M. D., & Brodzik, S. R. (2015). The variable nature of convection in the tropics and subtropics: A legacy of 16-years of the tropical rainfall measuring mission satellite. *Reviews of Geophysics*, *53*(3), 994-1021. doi: 10.1002/2015RG000488

Houze, Schmid, W., Fovell, R., & Schiesser, H. (1993). Hailstorms in switzerland: Left movers, right movers, and false hooks. *Monthly weather review*, *121*(12), 3345–3370.

Houze, R. A., & Betts, A. K. (1981). Convection in gate. *Reviews of Geophysics*, *19*(4), 541-576. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/RG019i004p00541` doi: https://doi.org/10.1029/RG019i004p00541

Huang, Hu, C., Huang, X., Chu, Y., Tseng, Y.-h., Zhang, G. J., & Lin, Y. (2018). A long-term tropical mesoscale convective systems dataset based on a novel objective automatic tracking algorithm. *Climate dynamics*, *51*(7-8), 3145–3159.

Huang, Wang, B., Wen, Z., & Chen, Z. (2021). Enhanced tropical eastern indian ocean rainfall breaks down the tropical easterly jet–indian rainfall relationship. *Journal of Climate*, *34*(8), 3039–3048.

Huffman, G., Stocker, E., Bolvin, D., Nelkin, E., & Tan, J. (2019). *Gpm imerg final precipitation l3 half hourly 0.1 degree x 0.1 degree v06, greenbelt, md, goddard earth sciences data and information services center (ges disc).*

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., ... others (2010). Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, *467*(7318), 951–954.

Kikuchi, K., & Wang, B. (2008, 06). Diurnal Precipitation Regimes in the Global Tropics*. *Journal of Climate*, *21*(11), 2680-2696. doi: 10.1175/2007JCLI2051.1

Kiladis, G. N., Thorncroft, C. D., & Hall, N. M. (2006). Three-dimensional structure and dynamics of african easterly waves. part i: Observations. *Journal of Atmospheric Sciences*, *63*(9), 2212–2230.

Kilpatrick, T. J., & Xie, S.-P. (2015). Ascat observations of downdrafts from mesoscale convective systems. *Geophysical Research Letters*, *42*(6), 1951-1958. doi: 10.1002/2015GL063025

Kim, D.-W., Protopapas, P., Bailer-Jones, C. A., Byun, Y.-I., Chang, S.-W., Marquette, J.-B., & Shin, M.-S. (2014). The epoch project-i. periodic variable stars in the eros-2 lmc database. *Astronomy & Astrophysics*, *566*, A43.

Kingsmill, D. E. (1995). Convection initiation associated with a sea-breeze front, a gust front, and their collision. *Monthly weather review*, *123*(10), 2913–2933.

Klocke, D., Brueck, M., Hohenegger, C., & Stevens, B. (2017). Rediscovery of the doldrums in storm-resolving simulations over the tropical atlantic. *Nature Geoscience*, *10*(12), 891–896.

Knupp, K. (2006). Observational analysis of a gust front to bore to solitary wave transition within an evolving nocturnal boundary layer. *Journal of the atmospheric sciences*, *63*(8), 2016–2035.

Koteswaram, P., & George, C. (1958). On the formation of monsoon depressions in the bay of bengal. *Indian Journal of Meteorology and Geophysics*, *9*, 9–22.

Krishnan, R., Zhang, C., & Sugi, M. (2000). Dynamics of breaks in the indian summer monsoon. *Journal of Atmospheric Sciences*, *57*(9), 1354–1372.

Lang, T. J. (2017). Investigating the seasonal and diurnal cycles of ocean vector winds near the philippines using rapidscat and ccmp. *Journal of Geophysical Research: Atmospheres*, *122*(18), 9668-9684. doi: 10.1002/2017JD027516

Lawrence, D. M., & Webster, P. J. (2002). The boreal summer intraseasonal oscillation: Relationship between northward and eastward movement of convection. *Journal of the atmospheric sciences*, *59*(9), 1593–1606.

Lepore, C., Tippett, M. K., & Allen, J. T. (2018). Cfsv2 monthly forecasts of tornado and hail activity. *Weather and Forecasting*, *33*(5), 1283–1297.

Lepore, C., Veneziano, D., & Molini, A. (2015). Temperature and cape dependence of rainfall extremes in the eastern united states. *Geophysical Research Letters*, *42*(1), 74–83.

Lin, Lee, M.-I., Kim, D., Kang, I.-S., & Frierson, D. M. (2008). The impacts of convective parameterization and moisture triggering on agcm-simulated convectively coupled equatorial waves. *Journal of Climate*, *21*(5), 883–909.

Lin, & Neelin, J. D. (2003). Toward stochastic deep convective parameterization in general circulation models. *Geophysical research letters*, *30*(4).

Liu, Y., Guo, L., Wu, G., & Wang, Z. (2010). Sensitivity of itcz configuration to cumulus convective parameterizations on an aqua planet. *Climate dynamics*, *34*(2-3), 223–240.

Long, J. A., Lawrence, R. L., Greenwood, M. C., Marshall, L., & Miller, P. R. (2013). Object-oriented crop classification using multitemporal etm+ slc-off imagery and random forest. *GIScience & Remote Sensing*, *50*(4), 418–436.

L'Heureux, M. L., Takahashi, K., Watkins, A. B., Barnston, A. G., Becker, E. J., Di Liberto, T. E., ... others (2017). Observing and predicting the 2015/16 el niño. *Bulletin of the American Meteorological Society*, *98*(7), 1363–1382.

Maddox, R. A., Chappell, C. F., & Hoxit, L. R. (1979). Synoptic and meso-α scale aspects of flash flood events. *Bulletin of the American Meteorological Society*, *60*(2), 115–123.

Manabe, S., Smagorinsky, J., & Strickler, R. F. (1965). Simulated climatology of a general circulation model with a hydrologic cycle. *Monthly Weather Review*, *93*(12), 769–798.

Mapes, B., Milliff, R., & Morzel, J. (2009). Composite life cycle of maritime tropical mesoscale convective systems in scatterometer and microwave satellite observations. *Journal of the atmospheric sciences*, *66*(1), 199–208.

Mapes, B., Tulich, S., Lin, J., & Zuidema, P. (2006). The mesoscale convection life cycle: Building block or prototype for large-scale tropical waves? *Dynamics of atmospheres and oceans*, *42*(1-4), 3–29.

Maxwell, A., Strager, M., Warner, T., Zégre, N., & Yuill, C. (2014). Comparison of naip orthophotography and rapideye satellite imagery for mapping of mining and mine reclamation. *GIScience & Remote Sensing*, *51*(3), 301–320.

Mekonnen, A., & Rossow, W. B. (2011). The interaction between deep convection and easterly waves over tropical north africa: A weather state perspective. *Journal of Climate*, *24*(16), 4276–4294.

Mlawer, Clough, S., & Kato, S. (1998). Shortwave clear-sky model measurement intercomparison using rrtm. In *Proceedings of the eighth arm science team meeting* (pp. 23–27).

Mlawer, Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: Rrtm, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, *102*(D14), 16663–16682.

Molinari, J., Knight, D., Dickinson, M., Vollaro, D., & Skubis, S. (1997). Potential vorticity, easterly waves, and eastern pacific tropical cyclogenesis. *Monthly weather review*, *125*(10), 2699–2708.

Molinari, J., Vollaro, D., Skubis, S., & Dickinson, M. (2000). Origins and mechanisms of eastern pacific tropical cyclogenesis: A case study. *Monthly weather review*, *128*(1), 125–139.

Mori, S., Jun-Ichi, H., Tauhid, Y. I., Yamanaka, M. D., Okamoto, N., Murata, F., ... Sribimawati, T. (2004). Diurnal land–sea rainfall peak migration over sumatera island, indonesian maritime continent, observed by trmm satellite and intensive rawinsonde soundings. *Monthly Weather Review*, *132*(8), 2021–2039.

Nakamura, H., Sampe, T., Tanimoto, Y., & Shimpo, A. (2004). Observed associations among storm tracks, jet streams and midlatitude oceanic fronts. *Earth's Climate: The Ocean–Atmosphere Interaction, Geophys. Monogr*, *147*, 329–345.

Nasuno, T., Miura, H., Satoh, M., Noda, A. T., & Oouchi, K. (2009). Multi-scale organization of convection in a global numerical simulation of the december 2006 mjo event using explicit moist processes. *Journal of the Meteorological Society of Japan. Ser. II*, *87*(2), 335–345.

Nesbitt, S. W., Cifelli, R., & Rutledge, S. A. (2006, 10). Storm Morphology and Rainfall Characteristics of TRMM Precipitation Features. *Monthly Weather Review*, *134*(10), 2702-2721. doi: 10.1175/MWR3200.1

Nesbitt, S. W., & Zipser, E. J. (2003, 05). The Diurnal Cycle of Rainfall and Convective Intensity according to Three Years of TRMM Measurements. *Journal of Climate*, *16*(10), 1456-1475. doi: 10.1175/1520-0442(2003)016⟨1456:TDCORA⟩2.0.CO;2

O'Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, *10*(10), 2548-2563. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001351` doi: https://doi.org/10.1029/2018MS001351

Olson, D. A. (1985). The impact of existing boundaries on the usefulness of operational model qpf. In *Preprints, sixth conf. on hydrometeorology, indianapolis, in, amer. meteor. soc* (pp. 277–283).

Pattanaik, D., Sahai, A., Muralikrishna, R. P., Mandal, R., & Dey, A. (2020). Active-break transitions of monsoons over india as predicted by coupled model ensembles. *Pure and Applied Geophysics*, *177*, 4391–4422.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, *12*, 2825–2830.

Pei, S., Shinoda, T., Soloviev, A., & Lien, R.-C. (2018). Upper ocean response to the atmospheric cold pools associated with the madden-julian oscillation. *Geophysical Research Letters*, *45*(10), 5020-5029. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL077825` doi: https://doi.org/10.1029/2018GL077825

Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(3), e1301.

Ramamurthy, K. (1969). Monsoon of india: some aspects of the 'break'in the indian southwest monsoon during july and august. *Forecasting manual*, *1*, 1–57.

Raschendorfer, M. (2001). The new turbulence parameterization of lm. *COSMO newsletter*, *1*, 89–97.

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, *115*(39), 9684–9689.

Raymond, D. D., López-Carrillo, C., & Cavazos, L. L. (1998). Case-studies of developing east pacific easterly waves. *Quarterly Journal of the Royal Meteorological Society*, *124*(550), 2005–2034.

Raymond, D. J., Raga, G. B., Bretherton, C. S., Molinari, J., López-Carrillo, C., & Fuchs, Z. (2003, 09). Convective Forcing in the Intertropical Convergence Zone of the Eastern Pacific. *Journal of the Atmospheric Sciences*, *60*(17), 2064-2082. doi: 10.1175/1520-0469(2003)060⟨2064:CFITIC⟩2.0.CO;2

Reed, R. J., & Jaffe, K. D. (1981). Diurnal variation of summer convection over west africa and the tropical eastern atlantic during 1974 and 1978. *Monthly Weather Review*, *109*(12), 2527–2534.

Reed, R. J., Norquist, D. C., & Recker, E. E. (1977). The structure and properties of african wave disturbances as observed during phase iii of gate. *Monthly Weather Review*, *105*(3), 317–333.

Riehl, H. (1954). *Tropical meteorology* (Tech. Rep.). McGraw-Hill.

Robe, F. R., & Emanuel, K. A. (2001). The effect of vertical wind shear on radiative–convective equilibrium states. *Journal of the atmospheric sciences*, *58*(11), 1427–1445.

Roca, R., & Fiolleau, T. (2020). Extreme precipitation in the tropics is closely associated with long-lived convective systems. *Communications Earth & Environment*, *1*(1), 1–6.

Roja Raman, M., Jagannadha Rao, V., Venkat Ratnam, M., Rajeevan, M., Rao, S., Narayana Rao, D., & Prabhakara Rao, N. (2009). Characteristics of the tropical easterly jet: Long-term trends and their features during active and break monsoon phases. *Journal of Geophysical Research: Atmospheres*, *114*(D19).

Romps, D. M., Seeley, J. T., Vollaro, D., & Molinari, J. (2014). Projected increase in lightning strikes in the united states due to global warming. *Science*, *346*(6211), 851–854.

Rotunno, R., Klemp, J. B., & Weisman, M. L. (1988). A theory for strong, long-lived squall lines. *Journal of Atmospheric Sciences*, *45*(3), 463–485.

Rowe, A. K., & Houze, R. A. (2015). Cloud organization and growth during the transition from suppressed to active mjo conditions. *Journal of Geophysical Research: Atmospheres*, *120*(19), 10,324-10,350. doi: 10.1002/2014JD022948

Ruppert, J. H., & Johnson, R. H. (2016). On the cumulus diurnal cycle over the tropical warm pool. *Journal of Advances in Modeling Earth Systems*, *8*(2), 669-690. doi: 10.1002/2015MS000610

Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., & Düben, P. (2019). Global cloud-resolving models. *Current Climate Change Reports*, *5*(3), 172–184.

Schiro, K. A., & Neelin, J. D. (2019, 03). Deep Convective Organization, Moisture Vertical Structure, and Convective Transition Using Deep-Inflow Mixing. *Journal of the Atmospheric Sciences*, *76*(4), 965-987. doi: 10.1175/JAS-D-18-0122.1

Schiro, K. A., Neelin, J. D., Adams, D. K., & Lintner, B. R. (2016, 09). Deep Convection and Column Water Vapor over Tropical Land versus Tropical Ocean: A Comparison between the Amazon and the Tropical Western Pacific. *Journal of the Atmospheric Sciences*, *73*(10), 4043-4063. doi: 10.1175/JAS-D-16-0119.1

Schlemmer, L., & Hohenegger, C. (2014, 07). The Formation of Wider and Deeper Clouds as a Result of Cold-Pool Dynamics. *Journal of the Atmospheric Sciences*, *71*(8), 2842-2858. doi: 10.1175/JAS-D-13-0170.1

Schrodin, R., & Heise, E. (2001). *The multi-layer version of the dwd soil model terra_lm.* DWD.

Schulzweida, U., Kornblueh, L., & Quast, R. (2006). Cdo user's guide. *Climate data operators, Version*, *1*(6).

SeaPAC. (2015). *Rapidscat level 2b ocean wind vectors in 12.5km slice composites version 1.1 accessed on 16 october 2019.* NASA PO.DAAC. doi: 10.5067/RSX12-L2B11

Seeley, J. T., & Romps, D. M. (2015). Why does tropical convective available potential energy (cape) increase with warming? *Geophysical Research Letters*, *42*(23), 10–429.

Serra, Y. L., & Houze Jr, R. A. (2002). Observations of variability on synoptic timescales in the east pacific itcz. *Journal of the atmospheric sciences*, *59*(10), 1723–1743.

Simpson, J. (1969). A comparison between laboratory and atmospheric density currents. *Quarterly Journal of the Royal Meteorological Society*, *95*(406), 758–765.

Simpson, J., Mansfield, D., & Milford, J. (1977). Inland penetration of sea-breeze fronts. *Quarterly Journal of the Royal Meteorological Society*, *103*(435), 47–76.

Spencer, P. L., & Stensrud, D. J. (1998). Simulating flash flood events: Importance of the subgrid representation of convection. *Monthly weather review*, *126*(11), 2884–2912.

Stensrud, D. J., & Fritsch, J. M. (1994). Mesoscale convective systems in weakly forced large-scale environments. part iii: Numerical simulations and implications for operational forecasting. *Monthly weather review*, *122*(9), 2084–2104.

Stevens, B., ACQUISTAPACE, C., HANSEN, A., HEINZE, R., KLINGER, C., KLOCKE, D., … ZÄNGL, G. (2020). The added value of large-eddy and storm-resolving models for simulating clouds and precipitation. *Journal of the Meteorological Society of Japan. Ser. II*, *98*(2), 395-435. doi: 10.2151/jmsj.2020-021

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., … others (2019). Dyamond: The dynamics of the atmospheric general circulation modeled on non-hydrostatic domains. *Progress in Earth and Planetary Science*, *6*(1), 61.

Stumpf, A., & Kerle, N. (2011). Object-oriented mapping of landslides using random forests. *Remote sensing of environment*, *115*(10), 2564–2577.

Tang, X., Cai, Q., Fang, J., & Tan, Z.-M. (2019). Land–sea contrast in the diurnal variation of precipitation from landfalling tropical cyclones. *Journal of Geophysical Research: Atmospheres*, *124*(22), 12010–12021.

Tao, W.-K., Li, X., Khain, A., Matsui, T., Lang, S., & Simpson, J. (2007). Role of atmospheric aerosol concentration on deep convective precipitation: Cloud-resolving model simulations. *Journal of Geophysical Research: Atmospheres*, *112*(D24).

Trapp, R. J., & Woznicki, J. M. (2017). Convectively induced stabilizations and subsequent recovery with supercell thunderstorms during the mesoscale predictability experiment (mpex). *Monthly Weather Review*, *145*(5), 1739–1754.

Ukkonen, P., & Mäkelä, A. (2019). Evaluation of machine learning classifiers for predicting deep convection. *Journal of Advances in Modeling Earth Systems*, *11*(6), 1784-1802. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001561` doi: https://doi.org/10.1029/2018MS001561

Varble, A., Zipser, E. J., Fridlind, A. M., Zhu, P., Ackerman, A. S., Chaboureau, J.-P., … Shipway, B. (2014). Evaluation of cloud-resolving and limited area model intercomparison simulations using twp-ice observations: 1. deep convective updraft properties. *Journal of Geophysical Research: Atmospheres*, *119*(24), 13–891.

Vickers, D., & Mahrt, L. (2006). A solution for flux contamination by mesoscale motions with very weak turbulence. *Boundary-layer meteorology*, *118*(3), 431–447.

Vogel, R., Konow, H., Schulz, H., & Zuidema, P. (2021). A climatology of trade-wind cumulus cold pools and their link to mesoscale cloud organization. *Atmospheric Chemistry and Physics Discussions*, *2021*, 1–32. Retrieved from `https://acp.copernicus.org/preprints/acp-2021-420/` doi: 10.5194/acp-2021-420

Wakimoto, R. M. (1982). The life cycle of thunderstorm gust fronts as viewed with doppler radar and rawinsonde data. *Monthly weather review*, *110*(8), 1060–1082.

Webster, P. J., & Hoyos, C. (2004). Prediction of monsoon rainfall and river discharge on 15–30-day time scales. *Bulletin of the American Meteorological Society*, *85*(11),

1745–1766.
Wei, Y., Pu, Z., & Zhang, C. (2020). Diurnal cycle of precipitation over the maritime continent under modulation of mjo: Perspectives from cloud-permitting scale simulations. *Journal of Geophysical Research: Atmospheres*, *125*(13), e2020JD032529.
Weisman, M. L., Klemp, J. B., & Rotunno, R. (1988). Structure and evolution of numerically simulated squall lines. *Journal of Atmospheric Sciences*, *45*(14), 1990–2013.
Williams, E. R., Geotis, S., Renno, N., Rutledge, S., Rasmussen, E., & Rickenbach, T. (1992). A radar and electrical study of tropical "hot towers". *Journal of Atmospheric Sciences*, *49*(15), 1386–1395.
Wills, S. M., Cronin, M. F., & Zhang, D. (2021). Cold pools observed by uncrewed surface vehicles in the central and eastern tropical pacific. *Geophysical Research Letters*, *48*(10), e2021GL093373. (e2021GL093373 2021GL093373) doi: https://doi.org/10.1029/2021GL093373
Xie, S., Zhang, M., Boyle, J. S., Cederwall, R. T., Potter, G. L., & Lin, W. (2004). Impact of a revised convective triggering mechanism on community atmosphere model, version 2, simulations: Results from short-range weather forecasts. *Journal of Geophysical Research: Atmospheres*, *109*(D14). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JD004692` doi: https://doi.org/10.1029/2004JD004692
Xu, K.-M., & Randall, D. A. (1996). Explicit simulation of cumulus ensembles with the gate phase iii data: Comparison with observations. *Journal of Atmospheric Sciences*, *53*(24), 3710–3736.
Xue, M. (2000). Density currents in two-layer shear flows. *Quarterly Journal of the Royal Meteorological Society*, *126*(565), 1301–1320.
Yanai, M., Maruyama, T., Nitta, T., & Hayashi, Y. (1968). Power spectra of large-scale disturbances over the tropical pacific. *Journal of the Meteorological Society of Japan. Ser. II*, *46*(4), 308–323.
Yang, G.-Y., & Slingo, J. (2001, 04). The Diurnal Cycle in the Tropics. *Monthly Weather Review*, *129*(4), 784-801. doi: 10.1175/1520-0493(2001)129⟨0784:TDCITT⟩2.0.CO;2
Zängl, G., Reinert, D., Rípodas, P., & Baldauf, M. (2015). The icon (icosahedral non-hydrostatic) modelling framework of dwd and mpi-m: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, *141*(687), 563–579.
Zehnder, J. A., Powell, D. M., & Ropp, D. L. (1999). The interaction of easterly waves, orography, and the intertropical convergence zone in the genesis of eastern pacific tropical cyclones. *Monthly weather review*, *127*(7), 1566–1585.
Zuidema, P., Torri, G., Muller, C., & Chandra, A. (2017). A survey of precipitation-induced atmospheric cold pools over oceans and their interactions with the larger-scale environment. *Surveys in Geophysics*, *38*(6), 1283–1305.

Process	Scheme
Turbulent mixing in the boundary layer	Turbulent Kinetic Energy (TKE; Raschendorfer, 2001)
Cloud microphysics	Bulk microphysics scheme predicting cloud water, rainwater, cloud ice, snow, and graupel (Baldauf et al., 2011)
Surface	Interactive surface flux and soil model (Schrodin & Heise, 2001)
Radiative Transfer	Rapid Radiative Transfer Model (RRTM; Mlawer et al., 1997, 1998)
Diagnostic Fractional Cloud Cover	Box Probability Distribution Function

Table 1: Summary of parameterization schemes used in ICON model simulation from DYAMOND protocol (Hohenegger et al., 2020)

Region	Minimum Latitude	Minimum Longitude	Maximum Latitude	Maximum Longitude
West Pacific (WPAC)	-10.5	142.5	15.5	182.5
East Pacific (EPAC)	-10.5	192.5	15.5	252.5
Atlantic (AO)	-21.5	319.5	23.5	360.0
Indian (IO)	-14.5	52.5	17.5	92.5

Table 2: Minimum and Maximum Latitude (Degree) and Longitude (Degree) for each region box

Hyperparameter	N estimators	Min samples split	Min samples leaf	Max features	Max depth	Bootstrap
Value	400	2	1	3	80	False

Table 3: Hyperparameter configuration used to train RF algorithm

.

Region	Global Tropics	Deep Tropics	WPAC	EPAC	AO	IO
Number of TF gridpoints	49021	29450	3551	4679	4373	4549

Table 4: Number of TF grid points in each region used to calculate Pearson's correlation values.

Figure File.

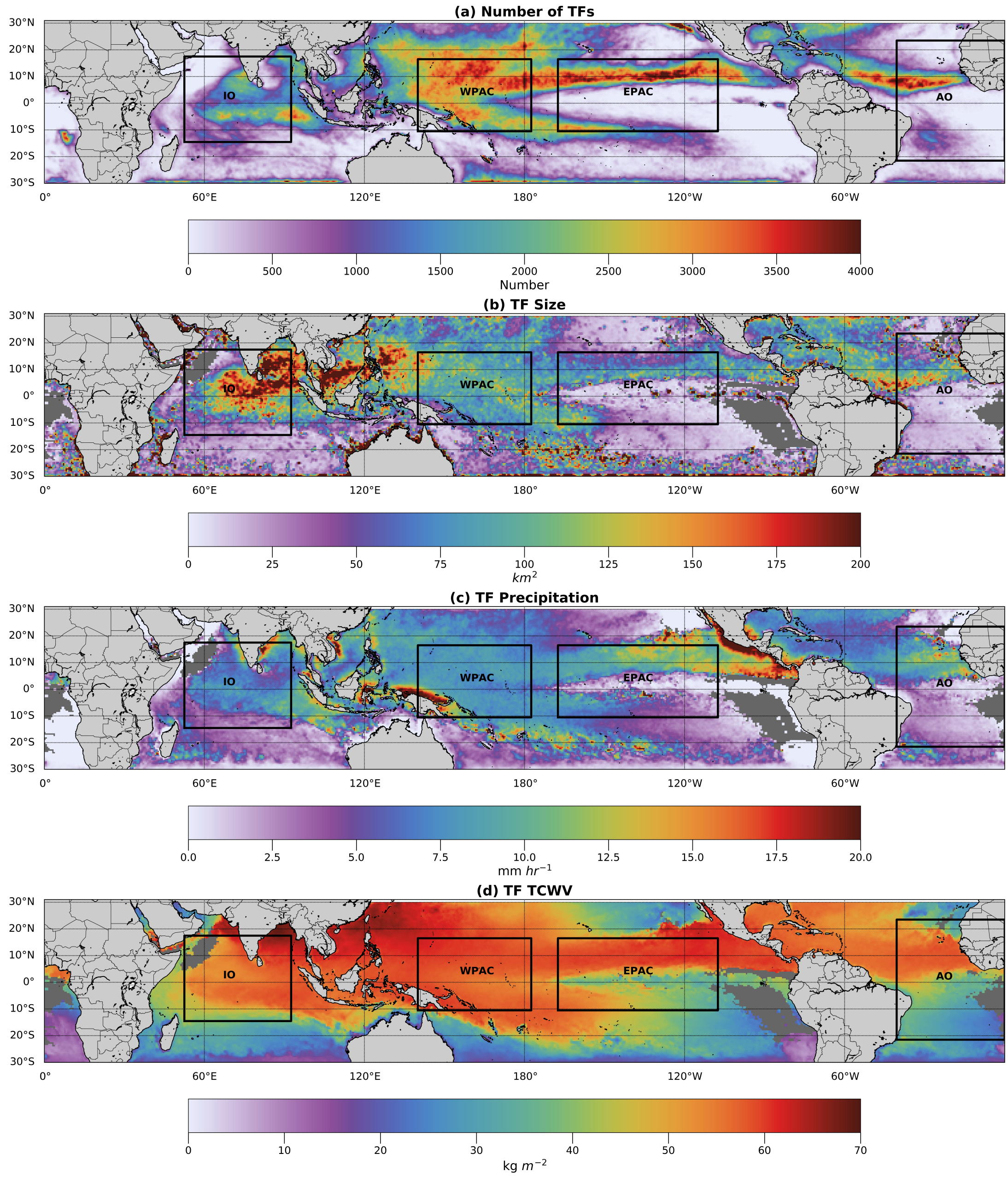


Figure 1: ICON-simulated (a) TF frequency, (b) TF size (km^2), (c) TF-attributed precipitation (mm hr^{-1}), (d) TF-attributed TCWV (kg m^{-2}) with WPAC, EPAC, AO, and IO boxes overlaid.

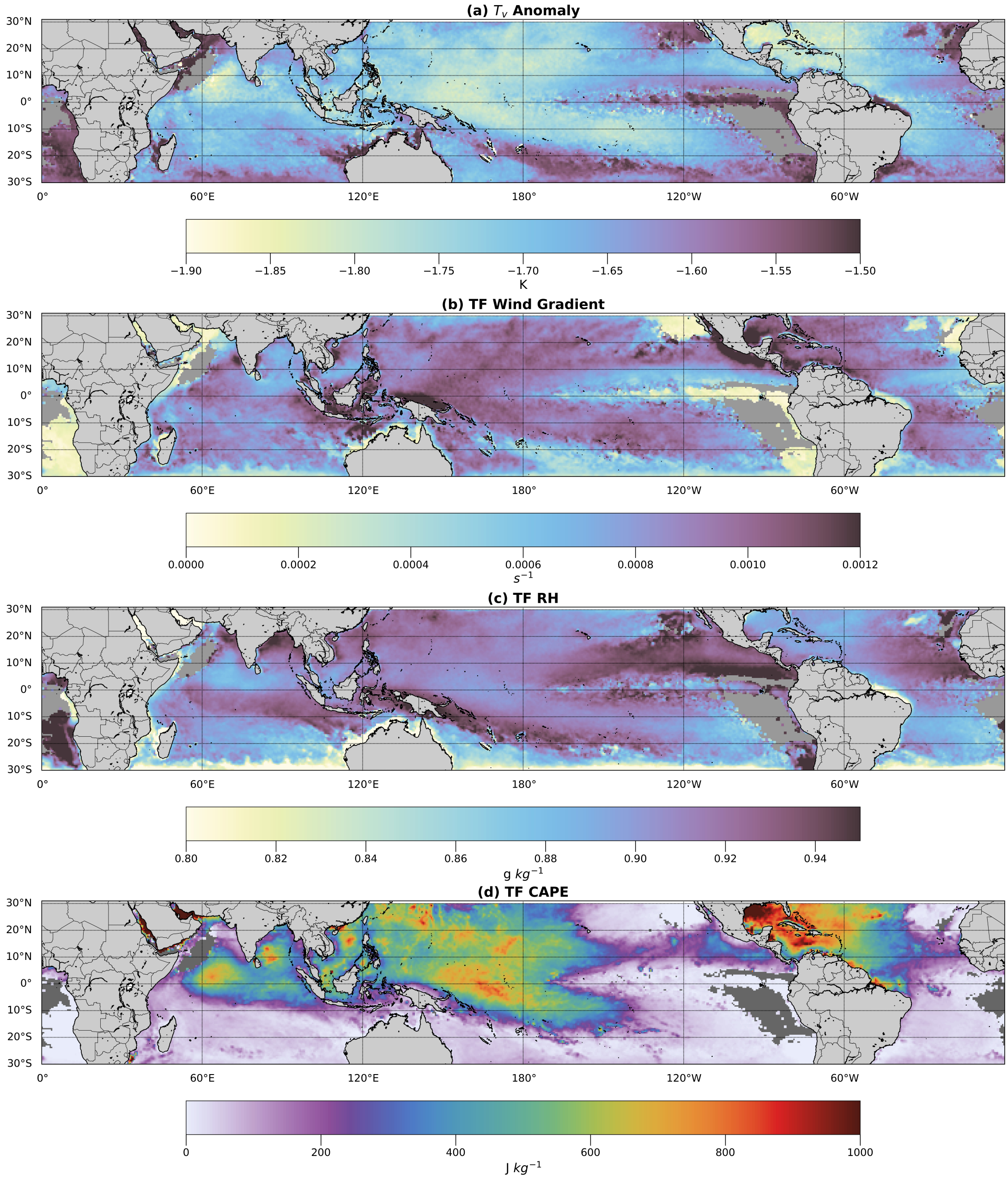


Figure 2: ICON-simulated mean (a) T_v anomaly (K), (b) TF wind gradient (s^{-1}), (c) TF-attributed RH (%), (d) TF-attributed CAPE (J kg^{-1})

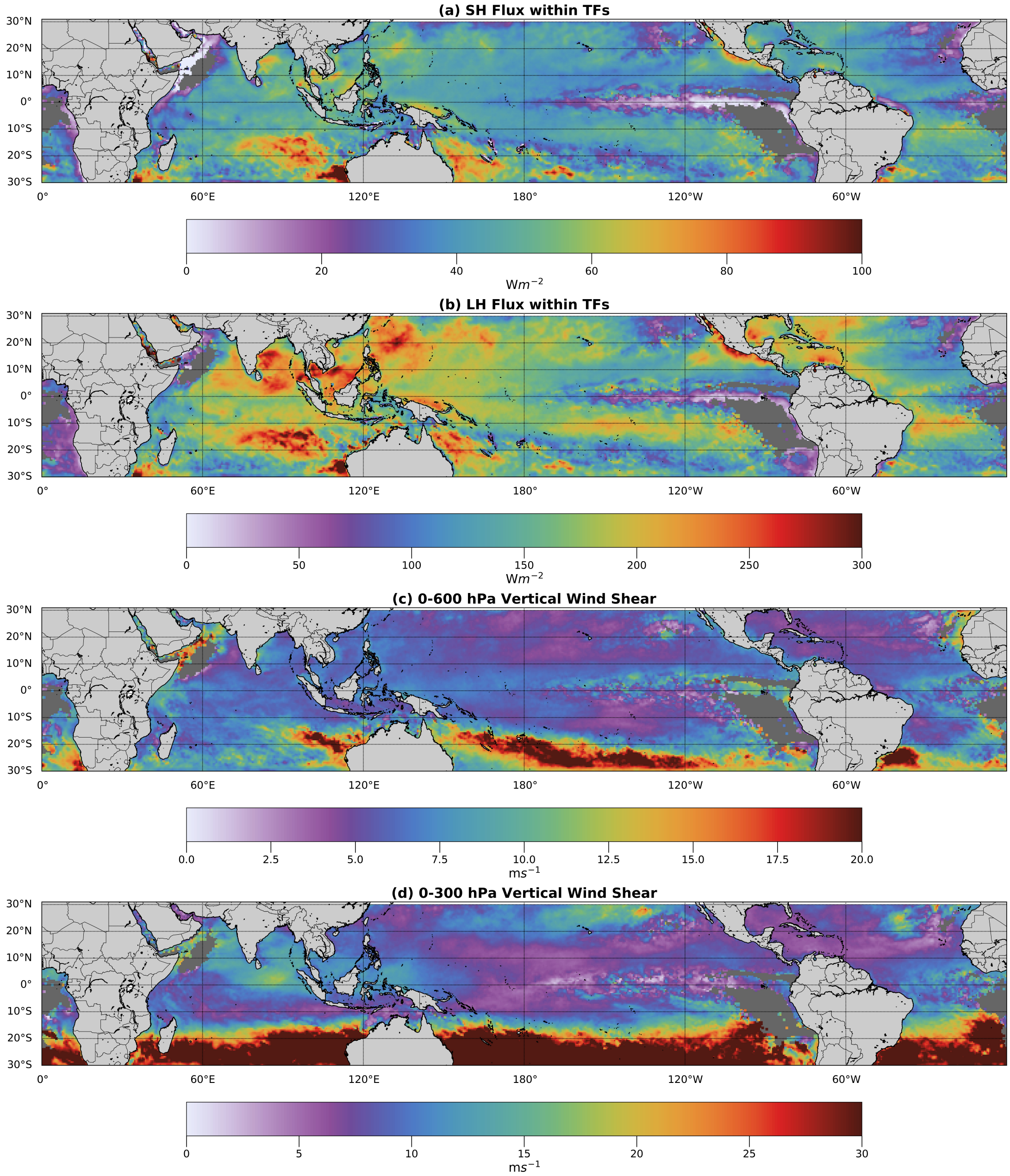


Figure 3: ICON-simulated mean (a) SHF (Wm^{-2}), (b) LHF (Wm^{-2}), (c) 0-600 hPa vertical wind shear (ms^{-1}), (d) 0-300 hPa vertical wind shear (ms^{-1})

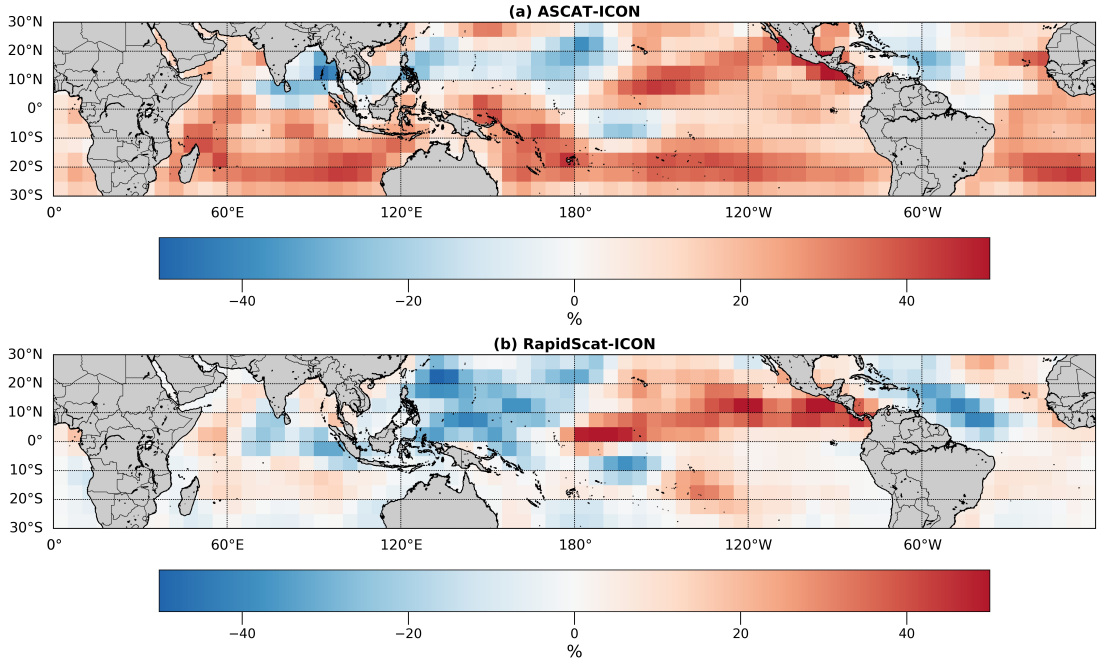


Figure 4: Percentage difference between (a) ASCAT GF (2007-2018) (b) RapidScat GF (2014-2016) and ICON TF (40-day) relative frequency Variation from Maximum for 01 Aug - 10 Sep 2016.

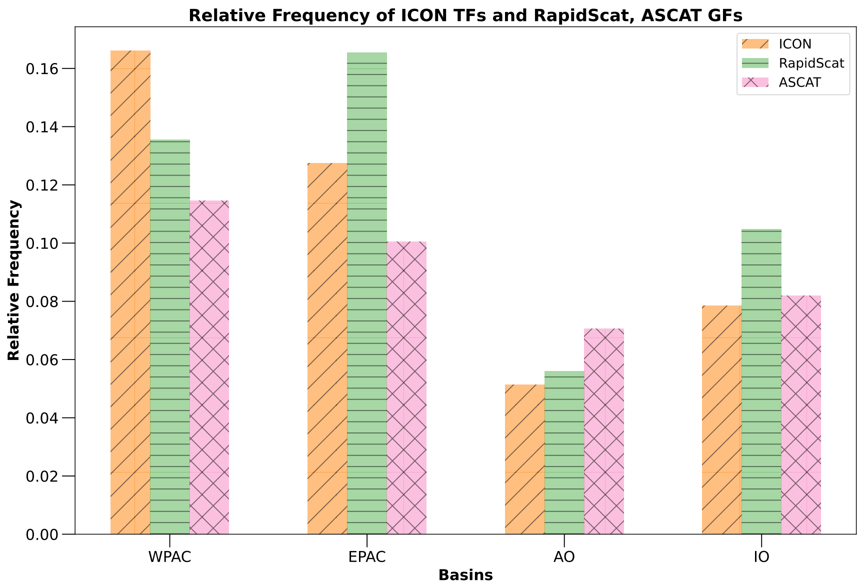


Figure 5: Relative frequency (basin/total) of ICON TFs with RapidScat and ASCAT GFs.

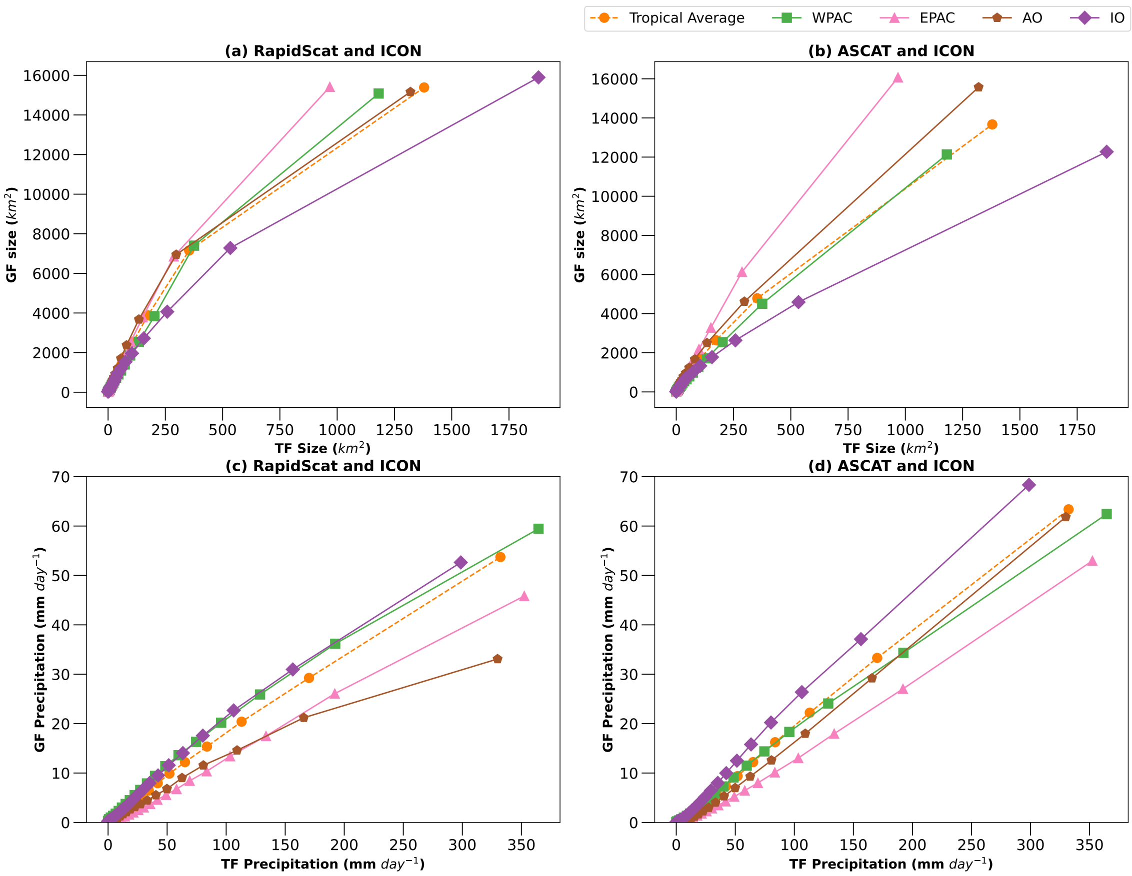


Figure 6: 5^{th} to 99^{th} Percentiles of ICON TF size(a)-(b) and precipitation (c)-(d) with (a),(c) RapidScat and (b),(d) ASCAT. The increment is 5% from 5^{th} to 95^{th} and 4% from 95^{th} to 99^{th} percentiles.

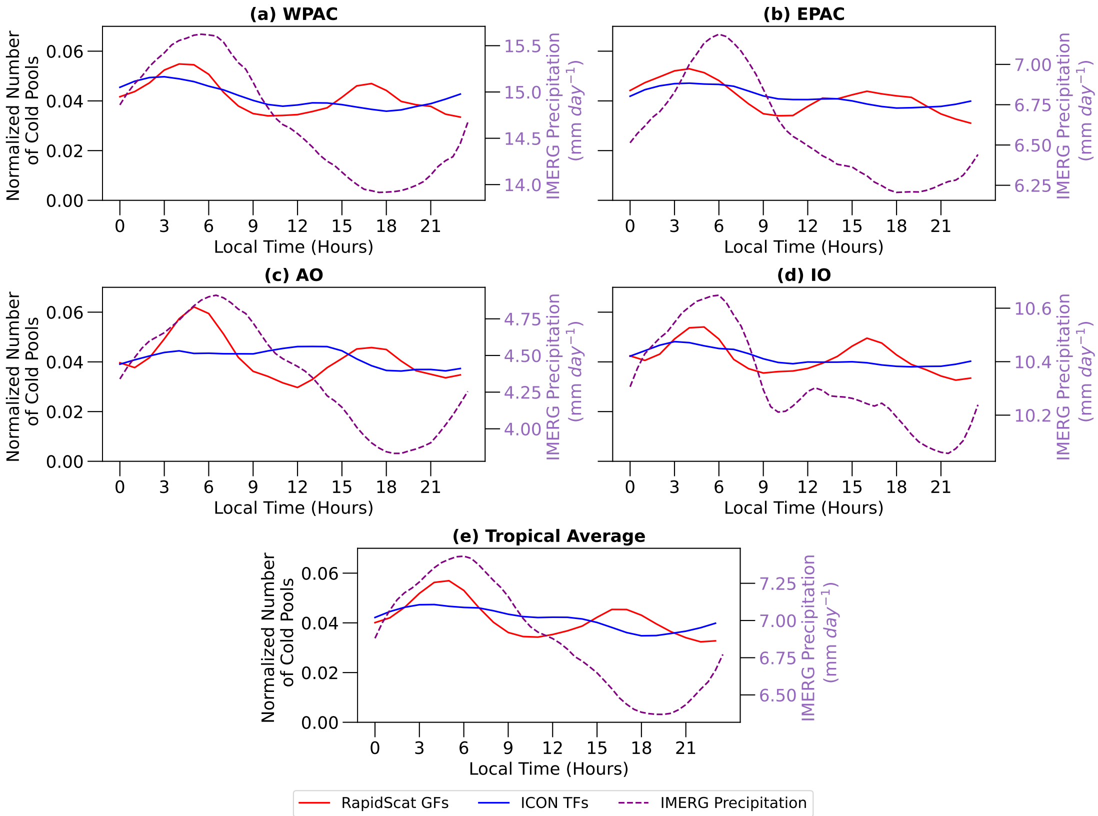


Figure 7: Diurnal cycle of ICON TF (solid blue), RapidScat GF (solid red) frequency, and IMERG precipitation (mm day^{-1}; dotted purple) for (a) WPAC, (b) EPAC, (c) AO, (d) IO, and (e) global tropical oceanic average.

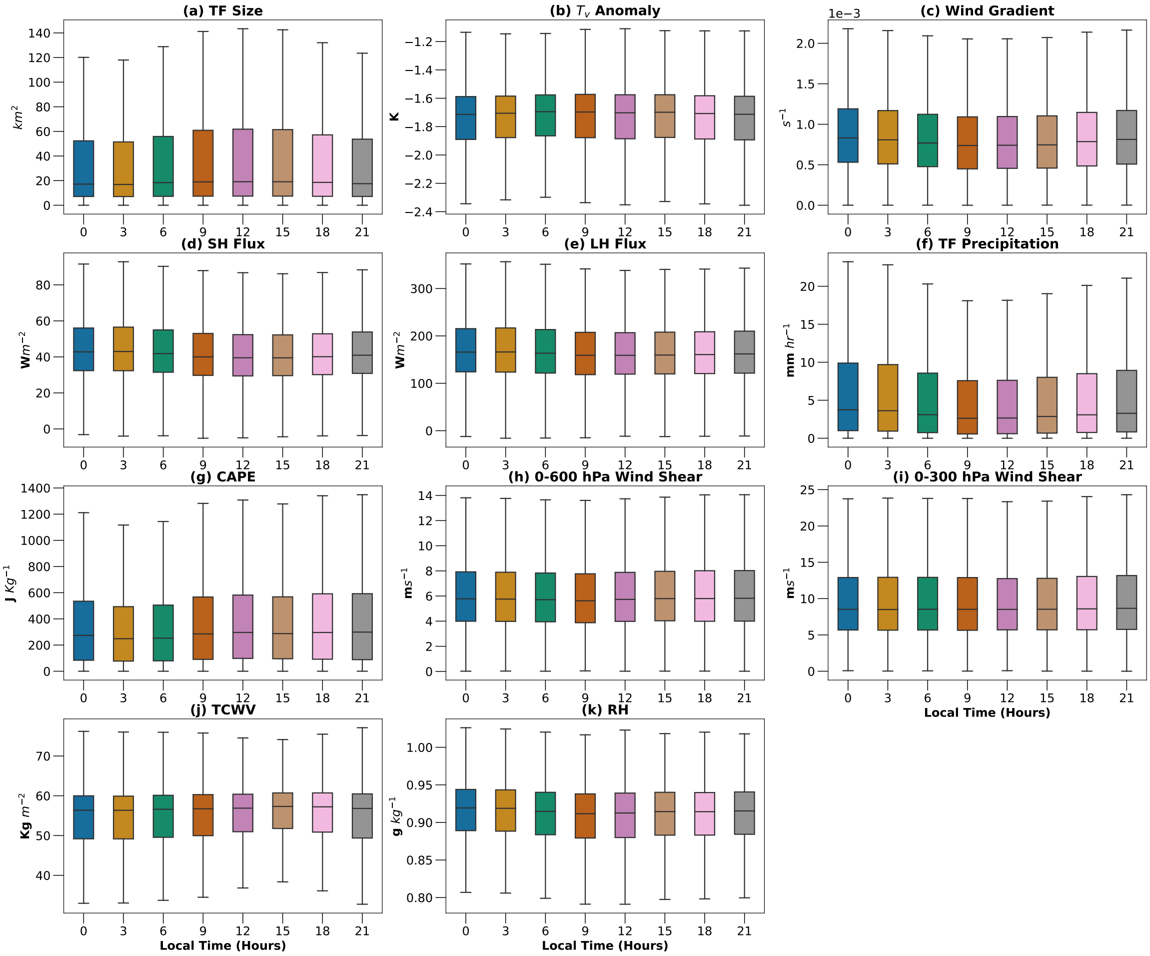


Figure 8: ICON-simulated diurnal cycle of (a) TF size (km^2), (b) T_v anomaly (K), (c) Wind gradient (s^{-1}), (d) SHF (Wm^{-2}), (e) LHF (Wm^{-2}), (f) TF precipitation (mm hr^{-1}), (g) CAPE (J kg^{-1}), (h) Surface - 600 hPa wind shear (ms^{-1}), (i) Surface - 300 hPa wind shear (ms^{-1}), (j) TCWV (kg m^{-2}), (k) RH (g kg^{-1})

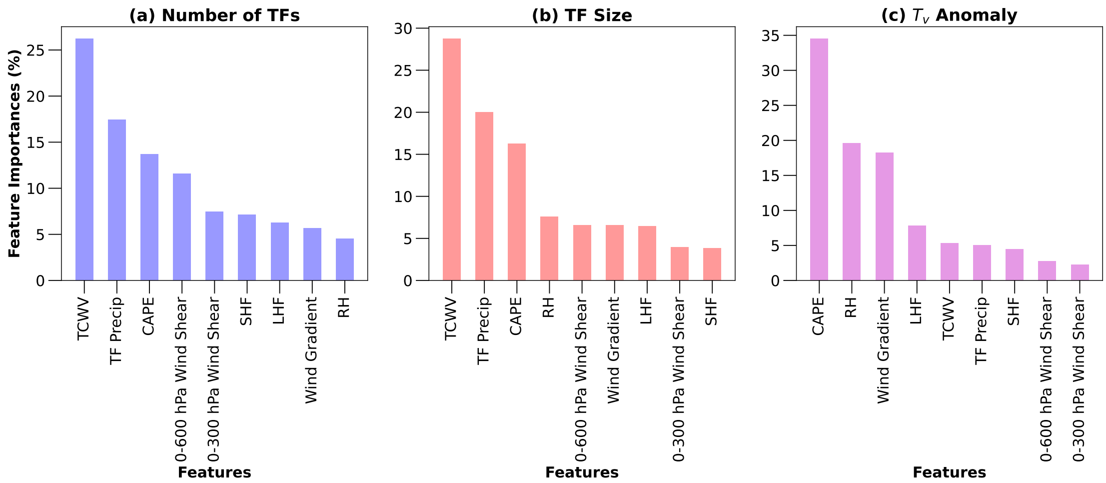


Figure 9: Relative feature importance for (a) TF frequency, (b) TF size (km^2), (c) TF intensity (T_v anomaly; K) obtained from RF algorithm.

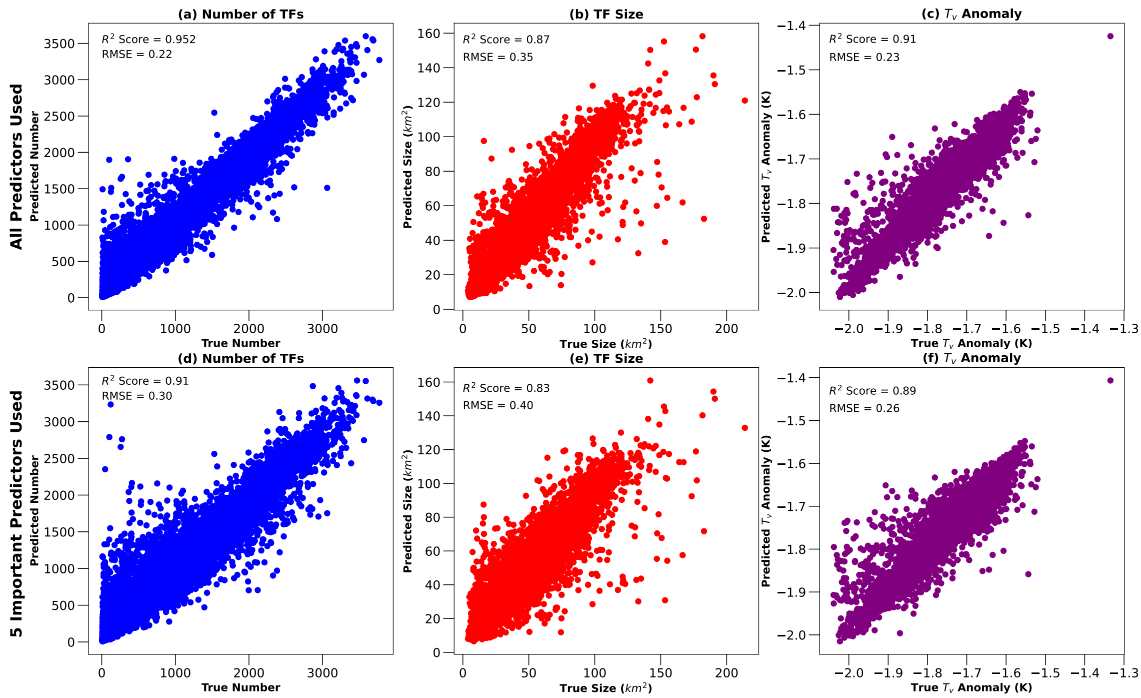


Figure 10: True (x-axis) versus predicted (y-axis) values using all the features (a)-(c) and the 5 most important features (d)-(f) for TF number frequency, size, and intensity. R^2 scores and RMSE are shown at the top left corner of each plot.

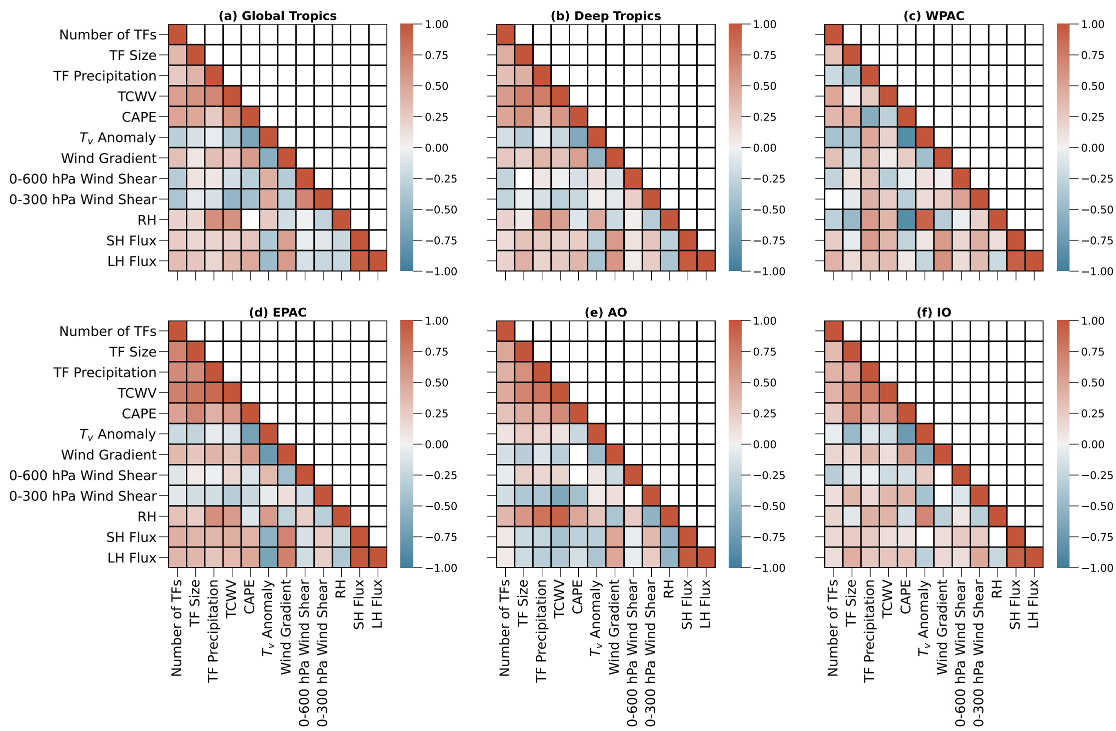


Figure 11: Pearson's correlation matrix for (a) global tropics, (b) deep tropics, (c) WPAC, (d) EPAC, (e) AO, and (f) IO. White boxes signify no statistically significant correlation.

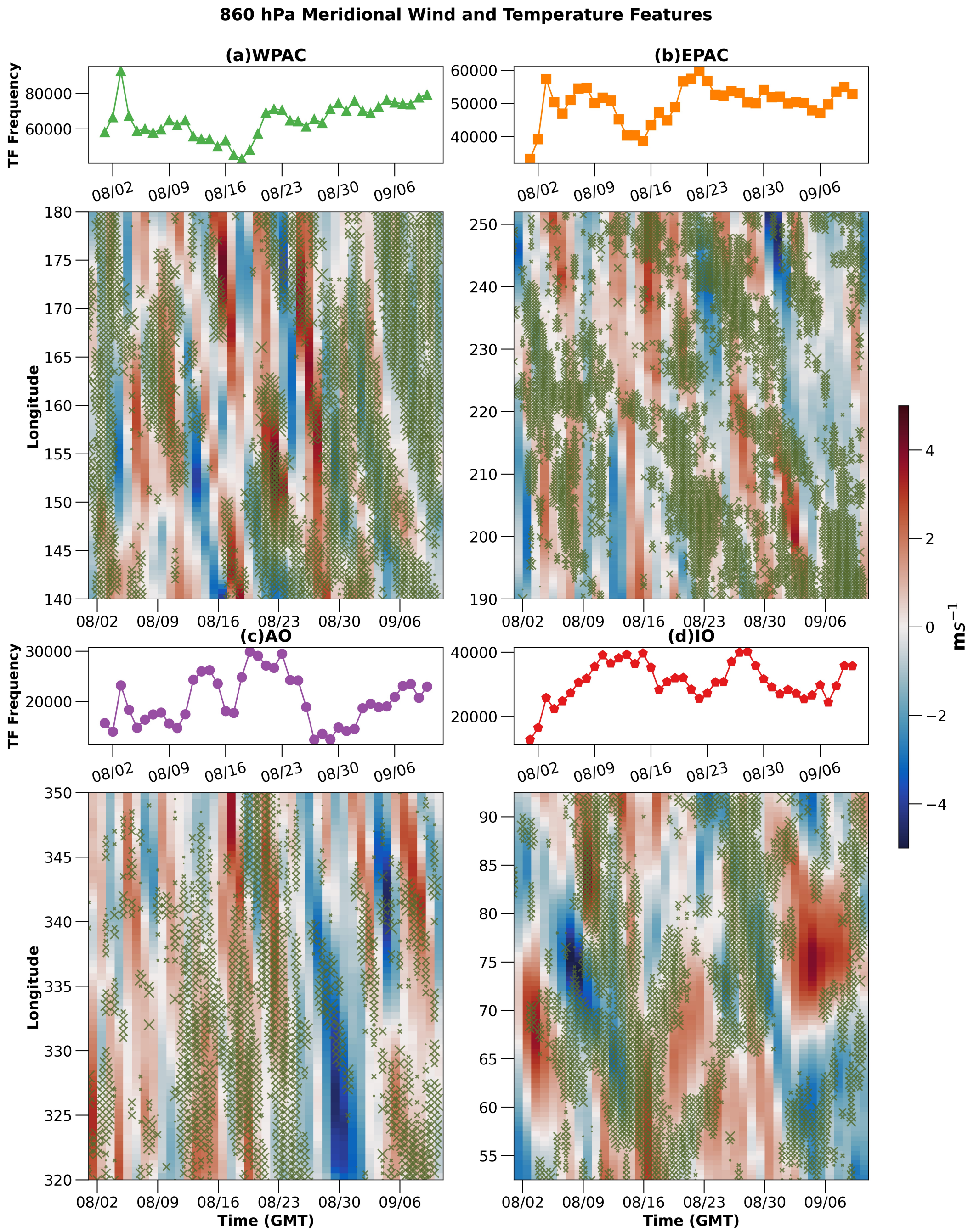
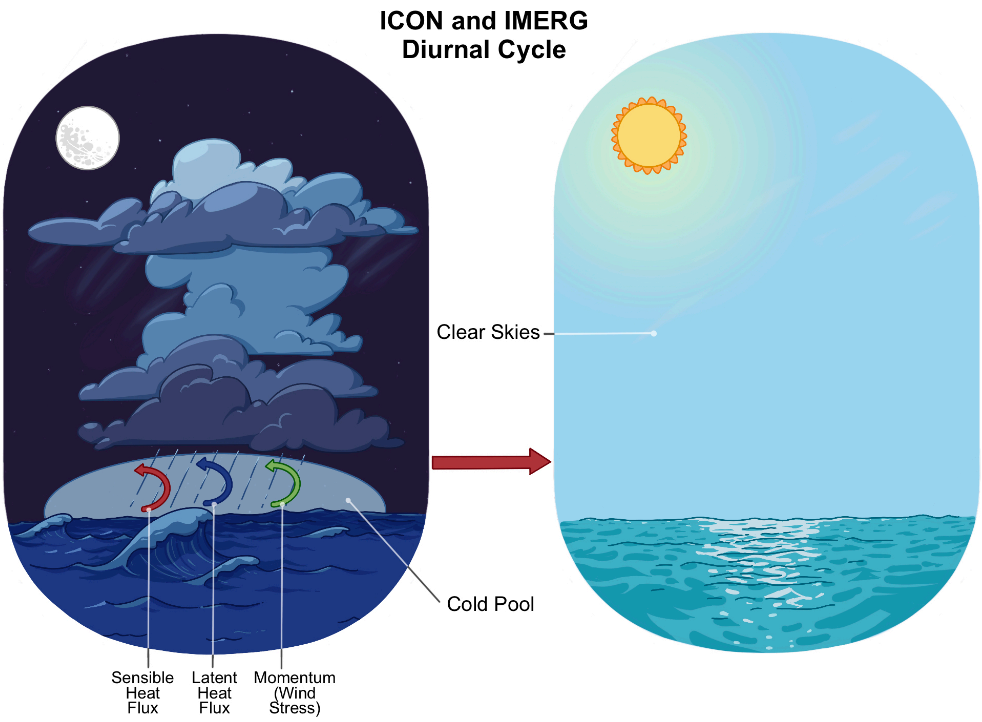
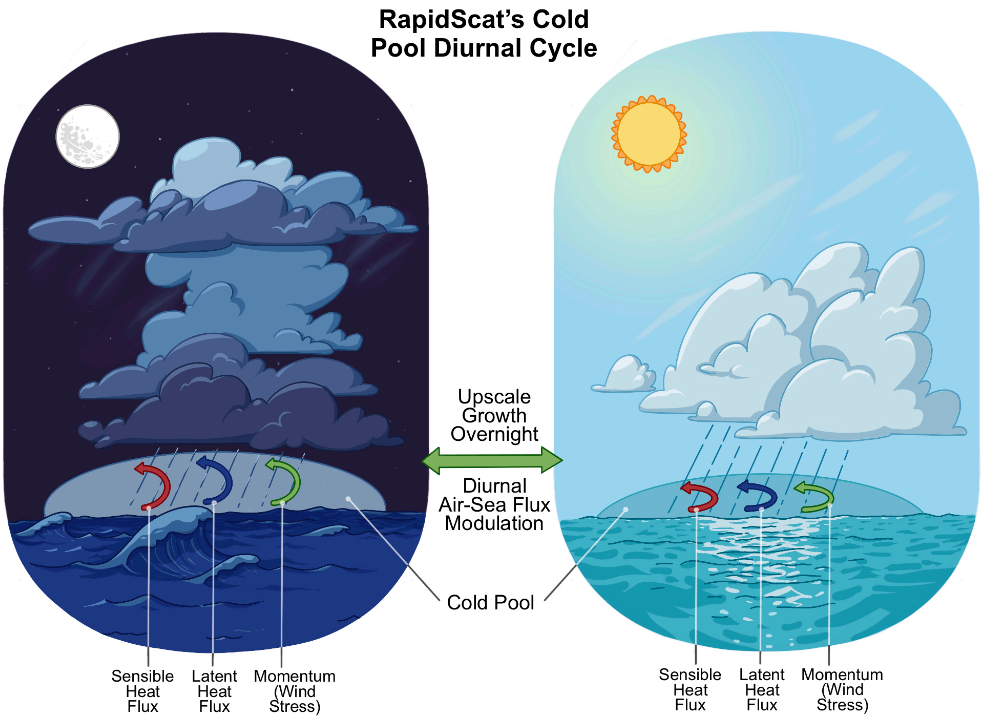


Figure 12: Time series of TF frequency and Hovmöller diagram between time and longitude showing TF number frequency anomaly (green markers) with 860 hPa meridional wind anomaly and time series of TF frequency for (a) WPAC, (b) EPAC, (c) AO, and (d) IO.



(a) ICON and IMERG



(b) RapidScat

Figure 13: Illustration depicting differences between RapidScat, ICON, and IMERG cold pool-convection properties