Rainfall and Convection in ERA5 and MERRA-2 over the Northern Equatorial Western Pacific during PISTON

YOLANDE L. SERRA,a STEVEN A. RUTLEDGE,b KYLE CHUDLER,b AND CHIDONG ZHANGc

a Cooperative Institute for Climate, Ocean, and Ecosystem Studies, University of Washington, Seattle, Washington
b Colorado State University, Fort Collins, Colorado
c Pacific Marine Environmental Laboratory, NOAA, Seattle, Washington

(Manuscript received 29 March 2022, in final form 4 October 2022)

ABSTRACT: This study evaluates rainfall, cloudiness, and related fields in the European Centre for Medium-Range Weather Forecasts fifth-generation climate reanalysis (ERA5) and the National Aeronautics and Space Administration’s Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2), gridded global reanalysis products against observations from the Office of Naval Research’s Propagation of Intraseasonal Tropical Oscillations (PISTON) field campaign. We focus on the first PISTON cruise, which took place from August to October 2018 in the northern equatorial western Pacific Ocean. We find biases in the mean surface heat and radiative fluxes consistent with observed biases in high and low cloud fraction and convective activity in the reanalyses. Biases in the high, middle, and low cloud fraction are also consistent with the biases in the thermodynamic profiles, with positive biases in upper-level humidity associated with excessive high cloud in both products, whereas negative biases in humidity above the boundary layer are associated with too few low and middle clouds and increased static stability. ERA5 exhibits a profile that is more top-heavy than that of MERRA-2 during periods dominated by MCSs and stronger upward motion during rainy periods, consistent with higher total rainfall in this product during PISTON. The coarser grid size in MERRA-2 relative to ERA5 and the fact that MERRA-2 did not assimilate PISTON data likely both contribute to the overall larger biases seen in MERRA-2. The observed biases in the reanalyses during PISTON have also been seen in comparisons of these products with satellite data, suggesting that the results of this study are more broadly applicable.

KEYWORDS: Radars/radar observations; Radiosonde/rawinsonde observations; Ship observations; Cumulus clouds; Model evaluation/performance; Reanalysis data

1. Introduction

Global gridded reanalysis products provide information about tropical weather and climate widely used for monitoring and improving understanding of the tropics and their global teleconnections. The European Centre for Medium-Range Weather Forecasts fifth-generation climate reanalysis (ERA5; Hersbach et al. 2020) and the National Aeronautics and Space Administration’s Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al. 2017) are two of the latest global gridded products to provide estimates of the state of the atmosphere at subdaily time scales and submesoscale (≤50 km) spatial scales. ERA5 replaced ERA-Interim in mid-2019. Key improvements important to this study include the increase from 6-hourly to hourly time resolution and the increase from ~70-km resolution to 31-km horizontal resolution (Hersbach et al. 2020). MERRA-2 replaced MERRA in early 2016. While the spatial resolution of about 50 km remains roughly the same as that of MERRA, improvements to the amount and type of satellite data assimilated into the reanalysis have helped reduce biases in the MERRA-2 global water cycle (Gelaro et al. 2017). In addition, MERRA-2 adjusts model precipitation based on subdaily rainfall observations, with the largest impacts in the tropics (Gelaro et al. 2017).

This study evaluates these products over the northern equatorial western Pacific, a key region for convectively coupled modes of variability including the Madden–Julian oscillation (MJO) (e.g., Zhang 2013; Jiang et al. 2015), boreal summer intraseasonal oscillation (BSISO) (e.g., Wang and Xie 1997; Jiang et al. 2004), and convectively coupled equatorial waves (Kiladis et al. 2009), including easterly or tropical depression (TD) waves (Chudler and Rutledge 2021; Sobel et al. 2021). In addition, this region is characterized by a southwesterly monsoonal flow during boreal summer and a high frequency of tropical cyclones that exceeds any other region on Earth (Gray 1968).
Convection coupled to synoptic and intraseasonal modes of variability in this region and throughout the tropics tends to be organized into mesoscale convective systems (MCSs) that consist of both convective cells and stratiform regions (e.g., Houze 1989). MCSs have distinctive signatures on the environmental heating profiles in the atmosphere when compared with isolated deep convection (IC) or shallow convection (e.g., Hartmann et al. 1984; Houze 1989; Schumacher et al. 2004) that result in distinct impacts on the large-scale circulation (Schumacher et al. 2004). Convection over tropical oceans tends to develop on diurnal time scales, with isolated systems in the afternoon maturing into larger MCSs by early morning under favorable conditions (e.g., Chen and Houze 1997; Sui et al. 1997; Yang and Slingo 2001; Liu and Zipser 2008; Sakaeda et al. 2018). The interaction of tropical convection and the large-scale circulation has been widely explored using satellite, in situ, and reanalysis products (e.g., Zhang et al. 2004; Back and Bretherton 2006; Hagos et al. 2010; Huaman and Takahashi 2016; Huaman and Schumacher 2018; Huaman et al. 2022). Given the difficulty of directly observing latent heating or vertical velocity over the open ocean, reanalyses are an important tool for understanding the upscale impacts of tropical convection on the large-scale circulation. However biases in these products present challenges for such studies (e.g., Miao et al. 2019; Wolding et al. 2022).

We focus our evaluation on the August–October 2018 period, when shipboard radiosonde, surface meteorological, and C-band radar observations are available in the northern equatorial western Pacific as part of the Office of Naval Research’s Propagation of Intraseasonal Tropical Oscillations (PISTON) field campaign. The 2018 PISTON period included the passage of four typhoons (Jebi, Mangkhut, Trami, and Kong-Rey) in the region associated with several TD waves observed at the ship (Chudler and Rutledge 2021). Outside these events, the region experienced below-normal monsoon and intraseasonal (MJO and BSISO) activity (Sobel et al. 2021). Using the 2018 PISTON data, this study addresses two main questions: 1) How well do the gridded products represent the background state of the northern equatorial western Pacific during the 2018 PISTON campaign? 2) How well do these products capture the diurnal cycle of clouds and rainfall in this region? The outcome of this study will inform future investigations of the northern equatorial western Pacific using these products and, more generally, document gridded product performance at capturing convective development in the region in the absence of west Pacific monsoon, BSISO, or MJO forcing.

The paper is organized as follows, the data and methods are described in section 2, followed by an overview of the time evolution of rain events during PISTON in section 3, and a discussion of biases in the environmental profiles and convective parameters in section 4. Section 5 examines the diurnal cycle in cloudiness, rainfall, and related parameters. A discussion and concluding remarks are provided in section 6.

2. Data and methods
a. Observations

As part of the PISTON field campaign, shipboard high temporal resolution time series of the lower troposphere and detailed radar observations of cloud vertical and horizontal structure and evolution were collected over the northern equatorial western Pacific from 22 August to 12 October 2018, with a port call from 10 to 14 September 2018 in Palau (Fig. 1). This study makes use of a subset of these observations to assess ERA5 and MERRA-2 global gridded products.

The Colorado State University SEA-POL C-band radar (https://seapol.colostate.edu/deployment/) was deployed on the R/V Thomas G. Thompson (hereinafter TGT) during PISTON. SEA-POL is a polarimetric Doppler radar fully stabilized for shipboard operations (Rutledge et al. 2019a,b). SEA-POL data products from the 2018 cruise used in this study are on a 1-km grid in the horizontal spanning from 5 to 120 km from the ship obtained over time intervals of 7–15 min. Hourly time series of precipitation estimates based on gridded radar reflectivity and SEA-POL’s polarimetric variables (Rutledge et al. 2019a,b) from 5 to 7, 10, 50, and 120 km from the ship provide time–space-averaged rain rates for comparison with the reanalyses. These precipitation time series are calculated at the 2-km level. Also used for this study is the precipitation feature database from Chudler and Rutledge (2021). A precipitation “feature” is defined as an ellipse fitted to a contiguous area of SEA-POL reflectivity pixels with a reflectivity greater than 17 dBZ (developed for comparison with the NASA GPM satellite radar, which has a minimum detectable reflectivity of 17 dBZ). Once a feature is identified, attributes such as area, maximum horizontal dimension (“length”), mean rain rate, and echo-top height are assigned to that feature. Features were then further broken down into a classification of isolated (length < 20 km), sub-MCS (area < 2000 km²), or MCS (area > 2000 km²) based on their size. This study uses hourly isolated (IC) and MCS feature classifications to help isolate periods of primarily convective or stratiform cloud types.
Upper-air soundings were collected eight times daily from the TGT during the 2018 PISTON cruise. The PISTON sounding data were quality controlled at Colorado State University (CSU) following the method of Ciesielski et al. (2014) to generate a dataset at 5-hPa resolution in the vertical. In addition to the sounding data, shipborne surface observations of wind speed and direction, precipitation, air temperature, and humidity from the TGT, provided by NOAA's Physical Sciences Laboratory (PSL), are used to characterize the lower atmosphere during the PISTON cruise. The PSL data include eddy covariance and bulk atmosphere–ocean fluxes of heat, moisture, and momentum. The bulk latent and sensible heat fluxes are calculated using the COARE bulk flux algorithm 3.6 (Fairall et al. 1996a,b, 2003; Edson et al. 2013). Positive heat fluxes are defined as heating the ocean (downward). The shipboard data also include sea surface temperature observations from a sea snake (Curry et al. 2004), which provides temperature measurements close to the sea surface (~0.1-m depth), and an optical rain gauge (ORG) for measuring rainfall. The quality-controlled 10-min surface meteorological and flux datasets from PSL are interpolated to hourly for the analyses in this study.

PSL also processed the ceilometer data from the ship. These data include the median cloud base height (CBH) and cloud fraction estimates at hourly time resolution, with CBH defined as the median of the lowest cloud layer in the 15-s raw data and cloud fraction defined as the fractional count of 15-s cloudy and obscured pixels within the time interval. The upper detection limit of this instrument is 7 km or about 400 hPa, roughly inclusive of low and middle clouds in the reanalyses. If few or no clouds are present the ceilometer can report unrealistically high values of CBH (E. Thompson 2022, personal communication). To alleviate this issue, we do not use CBH from the ceilometer for cloud fractions below 5%.

NASA Tropical Rainfall Measuring Mission (TRMM) 3B42 daily 0.25° rainfall (Huffman et al. 2007) are used to provide an additional observation of daily rainfall with similar spatial sampling to the SEA-POL radar and reanalysis gridded products. The 3B42 product makes use of the TMPA (TRMM Multisatellite Precipitation Analysis) Algorithm, which includes a bias correction based on the TRMM Combined Instrument (TCI) precipitation estimate. The final TMPA product tends to underestimate daily average rainfall on monthly time scales by 2%–18% in the tropics when compared with in situ gauges on buoys and islands/atolls (Huffman et al. 2007). These data have been interpolated to the ship track using the ship’s mean daily position.

b. Gridded products

MERRA-2 surface meteorological variables are provided at hourly time steps on a 0.5° latitude × 0.625° longitude grid from the Goddard Earth Sciences Data and Information Services Center (GES DISC) (GMAO 2008a,b). MERRA-2 pressure level data are provided on the same spatial grid but at 3-hourly time steps (GMAO 2015a,b). ERA5 surface meteorological and pressure level data are available at hourly time steps on a grid with 0.25° × 0.25° resolution from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (Hersbach et al. 2018a,b).

All reanalysis parameters on pressure levels have been interpolated to the ship track and radiosonde launch times for direct comparisons with ship upper-air data. Reanalysis surface pressure and 2-m temperature and humidity are used together with the 1000-hPa value of the profile to obtain a 10-m temperature and humidity. The 10-m values are added as the first point in these profiles. Similarly, the ship 10-m air temperature and humidity are interpolated to the ship sounding times to obtain the first point for these profiles, replacing the first point at 15.5 m in the ship dataset. For difference plots, ship profiles above 10 m are fit to the reanalysis vertical pressure levels. Biases are defined as the reanalysis minus the observations in all calculations.

The PISTON soundings were uploaded to the Global Telecommunications System in near real time during the PISTON cruise with few exceptions. MERRA-2 data assimilation does not appear to include these data, based on availability of ship data in the area for the month of September (M. Bosilovich 2021, personal communication). Copernicus support (H. Hersbach 2021, personal communication) confirms that ERA5 did assimilate PISTON upper-air and surface meteorological data in 2018 when available.

Sampling errors due to differences in the spatial scale of the reanalyses with one another as well as the observations will impact direct comparisons between PISTON in situ observations and these data products, particularly for rainfall, which is highly variable in both space and time. The standard deviation of the mean in the parameters and their biases, where the mean is defined over the full PISTON period (51 days), provides an estimate of the variability in space and time and thus provides some measure of the sampling errors. The standard deviations for the subsets of PISTON data during rainy (23) and nonrainy (11) days (section 3a) have reduced degrees of freedom and thus likely underestimate sampling errors for these comparisons. Where possible the degree to which the biases found in the PISTON region agree with previous studies is also noted.

c. Convective parameters

The lifting condensation level (LCL), level of free convection (LFC), and low-level static stability are used to compare the convective environment during PISTON with the reanalyses. PSL provides the LCL and LFC based on a mixed layer (ML) parcel as part of the radiosonde dataset and those values are used in this study. Following the methods of Ciesielski et al. (2014), we use the lowest 50 hPa of the profile to calculate these same parameters for an ML parcel in the reanalyses. We define low-level static stability as the difference in potential temperature between the 850- and 1000-hPa levels, to specifically examine impacts of any temperature biases on convective development. These parameters are available at 3-hourly intervals in the observations, matching the radiosonde time series. Reanalysis data are interpolated to this same time frequency.
d. Cloud base height

CBH in ERA5 is defined as the level above the first model level for which the cloud fraction is greater than 1% and the condensate content is greater than $1 \times 10^{-6}$ kg kg$^{-1}$ (https://apps.ecmwf.int/codes/grib/param-db?id=228023; accessed 9 April 2021), while MERRA-2 defines CBH as the top of the atmospheric boundary layer (e.g., Wright et al. 2020). To compare this parameter quantitatively with the median ceilometer CBH from the TGT, we use the definition of CBH from the ERA5 documentation to calculate a value for the reanalyses based on the vertical profiles of cloud fraction and cloud liquid water available for both products. An important difference is that the ERA5 CBH is calculated on model levels and at hourly time resolution, while our estimate is calculated on the coarser pressure level data and using the 3-hourly profile time series. This results in a median CBH for ERA5 about 100 m higher than that provided by ECMWF. CBH from the ceilometer is also interpolated to 3-hourly to match the reanalysis profile time series. Median CBH from TGT hourly data is 650 m, while that for the 3-hourly data is 714 m, suggesting a sensitivity to the sampling interval similar to that seen for the ERA5 CBH.

3. Meteorological conditions during the 2018 PISTON cruise

This study evaluates rainfall, cloudiness, and related fields in the reanalyses associated primarily with the passage of westward-propagating synoptic-scale disturbances during PISTON 2018 (Chudler and Rutledge 2021; Sobel et al. 2021). The synoptic disturbances are difficult to separate from the passage of several typhoons, with four typhoons (Jebi, Mangkhut, Trami, and Kong Rey) active while the ship was on station. Using the IBTrACS database (Knapp et al. 2018, 2010), we find that the TGT data used for this study were collected when the ship was more than 200 km from a storm center of any magnitude. SEA-POL precipitation features indicate IC was seen most frequently within 120 km of the ship, occurring 91% of the time, with MCSs occurring only 13% of the time (Chudler and Rutledge 2021). Although MCSs compose a small fraction of the precipitation features during PISTON, Chudler and Rutledge (2021) show that MCSs dominated the rain volume observed by SEA-POL. Of these events, three were associated with typhoons in the area. However, the typhoon centers were still outside of the 200-km criterion for our analysis.

a. Rain events

Figure 2a shows daily rainfall rates and cumulative rainfall amount at the ship, and Fig. 2e compares the daily rainfall using boxplots. The notches on the boxes in Figs. 2e–h indicate the 95% confidence limits on the median value (center line), while the box itself represents the interquartile range (IQR: from the 25th to 75th percentiles) of the data. The whiskers (dashed lines) show the full range of the data not considered outliers and, finally, the outliers (values exceeding $1.5 \times$ IQR or $\sim 2.7 \sigma$, where $\sigma$ is the standard deviation, above/below the median) are shown as plus symbols.

Consecutive periods with daily rain rates of $\geq 5$ mm day$^{-1}$ are identified as rain events or “rainy days” (thick black horizontal lines), where the daily threshold can be exceeded by any of the three observed rainfall products (TGT ORG, SEA-POL 120 km, or TRMM). There are 23 days meeting this threshold. Names of typhoons are listed next to rain events that occurred when these systems were at their closest proximity to the ship. Based on Chudler and Rutledge (2021), all other rain events with daily rain rates of $\geq 5$ mm day$^{-1}$ were coincident with MCS rainfall in the SEA-POL rainfall database. These MCS associated rain events are identified by Chudler et al. (2021) as the convectively active phases of westward-propagating TD-waves, which dominated the subseasonal variability at the ship during PISTON 2018 (Chudler and Rutledge 2021; Sobel et al. 2021). We also define a “nonrainy” day category as days with rain accumulations of less than 1 mm in all three observed rainfall products. There are 11 days that meet this criterion during PISTON.

SEA-POL 120-km rain rates are averages over all 1 km × 1 km grids within 5–120 km of the radar (nonraining grids included). The spatial averaging for SEA-POL rain rates results in overall lower hourly values relative to the TGT ORG, which is a point measurement. TRMM daily rain rates and accumulations are similar to those of SEA-POL, and both are below those of the TGT ORG, suggesting similar space–time sampling for these products during PISTON. Considering the reanalyses, both ERA5 and MERRA-2 have more frequent rain than any of the observationally based products throughout the PISTON period, with only 2 and 3 of 51 cruise days being “nonrainy” in ERA5 and MERRA-2, respectively, contrasting with the 11 nonrainy days in the combined TRMM, SEA-POL, and TGT ORG measurements. Rain rates in ERA5 are overall larger than those in MERRA-2, resulting in total accumulations of 551 mm over the PISTON period, as compared with MERRA-2 total accumulations of 414 mm (Table 1). Observed rainfall accumulations varied by instrument, with totals of 417, 168, and 214 mm for the TGT ORG, SEA-POL 120 km, and TRMM, respectively. The results in Figs. 2a and 2e and Table 1 suggest that both ERA5 and MERRA-2 overestimate rainfall during PISTON; however, sampling errors are also likely contributing to the differences seen in both daily rain rate and accumulation shown here. That being said, positive biases in rainfall in the tropical western Pacific in both ERA5 and MERRA-2 have been reported elsewhere (Bosilovich et al. 2017; Hersbach et al. 2020). While both reanalyses indicate a larger amount of rain throughout the period, they are producing significantly more rain on rainy days than on nonrainy days, suggesting that convection is favored and suppressed at similar times in the models and observations.

b. Surface heat fluxes and SST

We compare the surface heat fluxes (Figs. 2b,c,f,g) and SST (Figs. 2d,h) to explore the air–sea interaction at the ship in the reanalyses during the period, where positive heat fluxes
are defined as heating the ocean. The reanalyses follow the observed latent heat fluxes (LHF) during PISTON reasonably well for periods of a day or longer, while reanalysis sensible heat fluxes (SHF), particularly ERA5, show biases on all time scales represented. For example, for the Jebi rain event both ERA5 and MERRA-2 SST is too warm over several days, consistent with an underestimate (less cooling) of the SHF during this time in both products. The opposite situation is seen for Kong-Rey, when SST in both reanalyses is too cold and ERA5 shows a large positive (more cooling) bias in SHF. Overall, ERA5 LHF bias is $-216$ W m$^{-2}$ with an RMS of $42$ W m$^{-2}$ while the MERRA-2 SHF bias is $-1$ W m$^{-2}$ with an RMS of $8$ W m$^{-2}$ (Figs. 2f,g). The negative surface flux biases indicate more surface cooling in the reanalyses relative to the observations. Recalculation of the LHF and SHF using the COARE bulk flux algorithm v3.6, the same version used for the PISTON bulk flux estimates shown in Figs. 2b and 2c, reduces the biases in MERRA-2, particularly for LHF, but produces similar or slightly larger biases in ERA5 (not shown). These results suggest that both parameter biases and choice of bulk flux algorithm play a role in producing the differences shown in Figs. 2b, 2c, 2f, and 2g.

Reanalyses ingest SST from observations, and therefore it is not a prognostic variable in these products. However, it is

| TABLE 1. Summary of total as well as rainy and nonrainy day rainfall accumulation along the ship track during PISTON 2018 based on the TGT ORG, SEA-POL 120 km rain-rate estimates, TRMM 3B42, ERA5, and MERRA-2. |
|----------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                      | TGT ORG         | SEA-POL 120 km  | TRMM            | ERA5            | MERRA-2         |
| Total (mm)           | 417             | 168             | 214             | 551             | 414             |
| Rainy days (mm)      | 395             | 144             | 199             | 325             | 251             |
| Nonrainy days (mm)   | 3               | 3               | 0               | 44              | 38              |

Fig. 2. (a) Daily rainfall and accumulation, (b) hourly SHF, (c) hourly LHF, and (d) hourly SST. Thick horizontal black lines indicate observed rain events with daily accumulation $\geq 5$ mm in any one of the three rainfall products (TGT ORG, SEA-POL 120 km, and TRMM). (e)-(h) Corresponding boxplots of quantities in (a)-(d) over the PISTON period; daily rain rates are used in (e). Abbreviations in (e)-(h) are defined as $\text{TG} = \text{TGT}$, $\text{SE} = \text{SEAPOL 120 km}$, $\text{TR} = \text{TRMM}$, $\text{E5} = \text{ERA5}$, and $\text{M2} = \text{MERRA-2}$. Gray area indicates when TGT was in port.
important for surface fluxes, which are calculated by the models, and the lower tropospheric stability, among other factors influencing the development of convection. While the SST used by the reanalyses follows the overall pattern at the ship (Fig. 2d) and shows good overall agreement (Fig. 2h), these SST products lack a strong diurnal cycle, most evident at the end of leg 2 (following the port call in Palau from 10 to 14 September) when little rain (Fig. 2a) and strong insolation (Fig. 3a) is observed. The baseline SST during this period in the reanalyses is higher than that at the ship because they miss the observed nighttime cooling. The nighttime SSTs can stabilize the real atmosphere while the lower boundary in the reanalyses may remain unstable from the lack of surface cooling, possibly contributing to excess rainfall or low clouds. As noted above, the SST in the reanalyses also lacks the strong upper ocean cooling event associated with Jebi, resulting in lower-than-observed surface fluxes during this rain event in these products, whereas the opposite occurs for the Kong-Rey event.

c. Surface radiation and cloudiness

Figure 3 shows downward solar radiation (SWD) and downward longwave radiation (LWD) at the surface, as well as total cloud fraction $F_c$ for both the observations and reanalyses. Because LWD and $F_c$ have high variability on hourly time scales, a 12-h running mean has been applied to these values in Figs. 3b and 3c to facilitate comparison with the observations. Figure 4 shows reanalysis $F_c$ for low, middle, and high cloud, also smoothed with a 12-h running mean. Low clouds in ERA5 are on pressure levels up to 0.8 of the surface pressure and middle clouds are on pressure levels from 0.8 to 0.45 of the surface pressure (https://apps.ecmwf.int/codes/grib/param-db?id=186, accessed 21 July 2021), or about 800 hPa and 800 to 450 hPa, respectively, for a surface pressure of 1000 hPa. MERRA-2 defines low clouds as below the 700 hPa level and high clouds as above the 400 hPa level, with middle clouds in between (M. Bosilovich 2021, personal communication). Table 2 provides a summary of cloud coverage for these same layers and for total cloud fraction, along with total cloud fraction for the TGT ceilometer. As noted above, the ceilometer total cloud fraction is limited to below 7 km (400 hPa) and is thus most comparable to the reanalysis middle and low cloud fraction coverage. However, as overlapping cloud layers exist within the reanalysis grid, the middle and low cloud layers from the reanalyses are not expected to sum to the cloud fraction observed by the ceilometer.

The rain events identified in Fig. 2a are evident as reductions in the SWD (Fig. 3a) and increases in $F_c$ (Fig. 3c) at the ship. Similar reductions in SWD and increases in total $F_c$ are also seen in the reanalyses at these times, indicating that the
reanalyses are simulating the cloud cover associated with these events, albeit with some variation expected due to the difference in spatial representation of these products when compared with the point measurement at the ship. Figure 4a further suggests that while high clouds are present most of the time in the reanalyses, coverage is elevated surrounding rain events. That being said, both MERRA-2 and, to a lesser extent, ERA5 are systematically biased low in SWD (Figs. 3a,d) and exhibit larger total Fc relative to the ship on most days.

Along with high cloud, middle and low clouds are also elevated during rain events (Figs. 4b,c), suggesting the trimodal distribution of clouds during convection expected in the tropics (Johnson et al. 1999) and existence of MCSs. However, MERRA-2 tends to underestimate LWD for most days (Fig. 3e), suggesting an underestimate of low cloud amount or an overestimate of CBH in this product. This is particularly evident during rain events when MERRA-2 LWD is at times 10 W m$^{-2}$ or more below ERA5 and TGT, suggesting that during these events MERRA-2 low cloud base is higher (colder) or optically thinner than what is observed. The lack of low cloud coverage in MERRA-2 as compared with ERA5 (Figs. 4c,f) further supports the interpretation that this product underestimates low cloud fraction. MERRA-2 also has generally fewer middle clouds than ERA5 (Figs. 4b,e); however, the middle cloud coverage often exceeds that of low clouds for this model.

4. Mean state

4.1 Thermodynamic profiles

Figures 5a–c show the mean thermodynamic state of the troposphere as documented by the TGT radiosondes during PISTON 2018. The mean mixed layer height, estimated as the highest level at which the TGT median virtual potential temperature is within one standard deviation ($\pm 0.2^\circ$) of its 10-m
value, is at about 400 m or 960 hPa (not shown). At this level, the mean relative humidity reaches its maximum value (Fig. 5b), while below this level specific humidity (Fig. 5a) and potential temperature (Fig. 5c) are relatively constant.

Figures 5d–f show biases in reanalysis specific humidity, relative humidity, and air temperature. Both products show a similar pattern with height for specific humidity, having negative biases below 700 hPa and positive biases above this level. The biases are larger in MERRA-2, likely in part because this product did not assimilate PISTON data. The nearly -2 g kg$^{-1}$ bias near 850 hPa is the largest absolute bias with height in MERRA-2, while the bias of 1 g kg$^{-1}$ at 450 hPa is the largest positive bias in this product, with moisture biases increasing to over 250% of the observed value at 250 hPa in MERRA-2. ERA5 has its largest absolute bias in specific humidity at 975 hPa at nearly -1 g kg$^{-1}$, with the magnitude of the bias being less than 0.5 g kg$^{-1}$ throughout the remainder of the profile, but with the relative bias being largest (173%) in the upper troposphere as seen with MERRA-2.

As with specific humidity $q$, the biases in relative humidity (RH) and temperature $T$ (Figs. 5e,f) are generally larger for MERRA-2 than for ERA5, again likely in part because MERRA-2 did not benefit from having assimilated PISTON data. The largest temperature biases in ERA5 are seen from the surface to 950 hPa, reaching almost -1°C. A bias of about -0.6°C is seen in MERRA-2 near 925 hPa above the observed mixed layer, with a bias of -1°C from 650 to 700 hPa, followed by a bias of 1°C at 350 hPa. In both ERA5 and MERRA-2 cloud fraction is highest in the upper troposphere where the relative humidity biases are maximum, reaching over 40% coverage near 200 hPa in both products.

Both ERA5 and MERRA-2 make use of critical RH values for determining large-scale cloud fractions (e.g., Wright et al.)
with subgrid sources and sinks of condensate associated with convection also contributing to model total cloud fractions (Forbes et al. 2011; Molod 2012). The critical RH values are based on satellite cloud fraction and RH profiles, for which cloud fraction increases with increasing RH, with lower or higher critical RH values, respectively resulting in more or fewer clouds for a given RH (Quaas 2012). Miao et al. (2019) found that MERRA-2 overestimates high cloud in the tropical western North Pacific, particularly in Northern Hemisphere summer, and underestimates middle and low cloud in this region throughout the year. These authors attribute the underestimate of middle and low clouds in MERRA-2 to high values of critical RH relative to satellite observations at these levels. Wright et al. (2020) pointed out that the higher critical RH value used in MERRA-2 relative to ERA5 at midlevels results in a higher frequency of high relative humidity values in MERRA-2 at this level than other reanalyses including ERA5, consistent with the RH biases seen in Fig. 5e. The dry bias in MERRA-2 in both specific and relative humidity near 850 hPa may also play a role in the lack of low cloud in this model (Figs. 4f and 5e), as vertical cloud development would encounter a dry free troposphere that would work to evaporate the clouds above the boundary layer. Critical RH in MERRA-2 shows good agreement with observations above 250 hPa at low latitudes (Miao et al. 2019), suggesting that positive biases in RH at upper levels like those seen in the PISTON region (Fig. 5e) may contribute to an overestimate of high cloud as noted in Miao et al. (2019).

Like Miao et al. (2019), Wright et al. (2020) revealed that MERRA-2 has more cloud water in the upper troposphere (above 500 hPa) in the tropics than do other reanalysis products including ERA5. This same study found that upper tropospheric moist static energy (MSE) biases in convective regions are higher in MERRA-2 than other reanalyses. These MSE biases were attributed to both excessive humidity, thought to be associated with excess detrainment of cloud water, and warm biases, thought to be related to more intense cloud radiative heating at anvil level (Wright et al. 2020). These results provide insight into the positive moisture and temperature biases in MERRA-2 at upper levels seen during PISTON (Figs. 5d,f).

### b. Convective instability

To examine the impact of the mean biases in the environment thermodynamic profiles on measures associated with the initiation of convection, we compare the LCL, LFC, 850–1000-hPa static stability, and CBH in the reanalyses with those from the ship soundings (Fig. 6). The observed LCL sits above the estimated mean mixed layer height of 415 m, with the reanalysis LCL being somewhat higher than the observed LCL. Biases in the ML and parcel environment affecting undilute adiabatic ascent result in a higher LFC for both gridded products, with the larger biases in ERA5 reflecting the larger ML biases in this product (Figs. 5d,f). CBH is close to the LCL in TGT and ERA5, while it is closer to the LFC in MERRA-2 reflecting the larger biases in MERRA-2 above the ML relative to ERA5 (Figs. 5d,f). The high bias in CBH for MERRA-2 seen here was predicted from the LWD comparisons with the ship (Figs. 3b,e). In addition, CBH is remarkably consistent in ERA5 as compared with TGT, suggesting a lack of variability in low cloud in ERA5, while MERRA-2 shows too much variability in CBH relative to the observations.

The lower tropospheric stability parameter indicates greater static stability in the reanalyses than in the observations, also consistent with the low-level air temperature biases in these products. The larger values in static stability for ERA5 than in MERRA-2 additionally reflect the larger negative bias in near-surface air temperature in ERA5 over MERRA-2 (Fig. 5f).

### c. Vertical structure of precipitating features

Sub-grid-scale latent heating is a source of upward motion in reanalyses, allowing reanalysis vertical velocity profiles to be used to evaluate the presence of shallow, deep, and stratiform convective cloud types in these models (Schumacher et al. 2004, 2007). In general, a high stratiform to deep convection ratio produces a vertical velocity profile that peaks above that of deep convection alone in the tropics (e.g., Schumacher et al. 2004). As MCS cloud features develop extensive stratiform regions in addition to deep convection (e.g., Houze 2004), an elevated peak in vertical velocity is expected for this category relative to that of regions dominated by predominantly

<table>
<thead>
<tr>
<th><img src="https://example.com/f6.png" alt="Fig. 6" /></th>
<th><img src="https://example.com/f6.png" alt="Fig. 6" /></th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) LCL [km]</td>
<td>(b) LFC [km]</td>
</tr>
<tr>
<td><img src="https://example.com/f6.png" alt="Fig. 6" /></td>
<td><img src="https://example.com/f6.png" alt="Fig. 6" /></td>
</tr>
<tr>
<td>(c) θ [850-1000 hPa]</td>
<td>(d) CBH [km]</td>
</tr>
</tbody>
</table>

### FIG. 6. Comparisons of TGT, ERA5 (E5), and MERRA-2 (M-2) (a) lifting condensation level (LCL), (b) level of free convection (LFC), (c) static stability parameter, and (d) cloud base height (CBH).
IC features, which produce more bottom-heavy vertical velocity profiles.

Figure 7 shows mean profiles of vertical velocity $\omega$ (solid lines) and cloud fraction $F_c$ (dashed lines) at the ship over the PISTON period for both ERA5 and MERRA-2. Error bars and shaded regions show 95% confidence limits estimated as 2 times the standard deviation of the mean. Parameters are averaged over rain rates at the time of the profile for rates (a) <0.05 and (b), (c) >0.3 mm h$^{-1}$. The profiles in (b) and (c) are also composited for SEA-POL radar MCS cloud feature areal coverage >50% and isolated cloud feature areal coverage >95%, respectively. The numbers in parentheses in the legends indicate the number of profiles that contribute to the composite for that product. Note the difference in vertical velocity scale in (b).

IC features, which produce more bottom-heavy vertical velocity profiles.

Figure 7 shows mean profiles of vertical velocity and cloud fraction for the reanalyses for nonrainy profiles (rain rates < 0.05 mm h$^{-1}$; Fig. 7a) and rainy profiles (rain rates > 0.3 mm h$^{-1}$) for periods when the SEA-POL radar indicates >50% MCS cloud feature coverage (Fig. 7b) and >95% IC feature coverage (Fig. 7c) around the ship. Rain rates are from each of the reanalysis products at the time of the profile. Percent coverage of MCS and IC features is calculated with respect to the sum of all features (i.e., IC + subMCS + MCS). The 0.05 mm h$^{-1}$ threshold is chosen for the nonrainy profiles as these products have no hourly periods of zero rain rate. The rain-rate threshold for rainy profiles was set together with the fraction of MCS or IC features to maximize the number of samples while also isolating the periods when one or the other precipitating feature type was dominant across the region.

For nonrainy profiles, MERRA-2 indicates subsiding motion throughout the troposphere below 13 km, while ERA5 indicates neutral to slightly upward motion. Both products indicate high cloud fractions of about 0.25–0.30 above 12 km for these periods. The subsidence below 3 km in MERRA-2 suggests greater radiative cooling is occurring in this product than in ERA5, overwhelming subgrid-scale upward motion associated with low cloud production. The peak in subsidence near 1 km
lies below the dry bias in MERRA-2 (Figs. 5d,e), which may contribute to greater radiative cooling at this level relative to ERA5. The peak subsidence at 11 km in MERRA-2 lies below the peak high cloud fraction at 14 km (Fig. 7a), where we might expect to see weak upward motion as in ERA5.

Figures 7b and 7c show ERA5 and MERRA-2 rainy profiles of vertical velocity and cloud fraction for periods when MCS and IC features exceed 50% and 95%, respectively, of the total number of precipitating features in the database. Overall, both products show upward motion throughout the troposphere during rainy periods, suggesting that grid-scale ascent is associated with the observed enhanced convection (stratiform or deep) represented by these composites. For MERRA-2 the shape of the vertical velocity profile is similar for the MCS and IC composites, but with a larger peak magnitude seen for the MCS composite throughout the profile. This contrasts the vertical velocity profiles in ERA5, where the MCS composite is more top-heavy than in the IC composite, consistent with subgrid-scale stratiform heating expected for these cases. In addition, MERRA-2 shows similar cloud fractions for the MCS and IC composites, while ERA5 shows larger cloud fractions at nearly all levels for the MCS composite relative to the IC composite. Below 3 km, a peak in upward motion is seen in both MCS and IC composites for both products suggesting the possible contribution from shallow convection during rainy periods (e.g., Schumacher et al. 2004).

As the model grids are larger than the IC precipitating features (<20 km), the weaker vertical velocities and lower cloud fractions at low and middle levels in the reanalyses, particularly in ERA5, relative to the MCS composites are consistent with the subgrid-scale convective activity of IC cases. However, the small middle cloud fractions in both reanalyses for the MCS profile is surprising given these systems are generally larger than a single grid in the models and more extensive clouds at midlevels would be expected. Overall, the vertical velocity and cloud fraction profiles shown in Fig. 7 and the cloud fraction comparisons in Fig. 4 and Table 2 suggest that ERA5 has a more trmodal distribution of cumulus and/or MCS-like clouds than MERRA-2 and better reproduces the expected top-heavy environmental heating profile for the MCS composite. The high cloud fractions and weaker upward motion above 12 km in both reanalyses further suggest that radiative heating of anvil clouds (e.g., Wright et al. 2020) is present during rainy periods.

5. Diurnal cycle in convection

The diurnal cycle is a prominent mode of short period variability in clouds and rainfall in the tropics (e.g., Chen and Houze 1997; Sui et al. 1997; Yang and Slingo 2001; Liu and Zipser 2008; Sakaeda et al. 2018). Cloud development from nonprecipitating to precipitating features reveals important information about the time scales and vertical extent of the processes contributing to the production of precipitation. A goal of this aspect of our study is to determine how well the reanalyses can capture this development on subdaily time scales. We separate nonrainy (all observed daily rainfall accumulations < 1 mm; 11 days) and rainy (at least one observed daily rainfall accumulation > 5 mm; 23 days) to isolate diurnal variability in clouds and related parameters on rain event days and contrast it with that on clear days. The rainy days in this section are the same as those discussed in section 2 and are based on observed daily rain, not the model rain rates used for the profiles in Fig. 7.

a. Clouds and precipitation

The diurnal cycle in observed rainfall and the cloud parameters measured at the ship are shown in Fig. 8. For this discussion we show the hourly-averaged TGT ORG and SEA-POL 120- and 10-km rain rates for all days, with the SEA-POL values scaled so that their 24-h mean values match those of the TGT ORG for improved readability (Fig. 8a). As discussed above, SEA-POL rain rates are averages of all 1-km grids within 5–120 km (SEA-POL 120 km) and 5–10 km (SEA-POL 10 km) of the radar (nonraining grid points included). The additional spatial averaging for SEA-POL rain rates results in overall lower hourly values than the TGT ORG. The 30- and 0-dBZ echo-top heights are also shown in Fig. 8b as a measure of the depth of the convective cells and cloud top, respectively. In fact, cloud top will extend beyond the 0-dBZ echo top, but the relative difference between the 0-dBZ echo top on nonrainy and rainy days is a helpful measure of the change in cloud vertical extent available from the ship observations.

While both IC and MCS counts are higher on rainy days, IC counts on nonrainy days (Fig. 8a) are considerably higher than MCS counts (Fig. 8b), suggesting that what little rain does fall on these days is primarily in the form of isolated convection of limited spatial extent and with convective cores below 5 km MSL on average (Fig. 8b). The nonrainy day IC counts are distributed throughout the day, while the few MCSs present occur primarily from 0000 to 0400 and from 0900 to 1300 LT. The 30-dBZ echo-top heights follow the diurnal cycle in MCS counts, peaking during 0000–0400 and 0900–1300 LT at around 4–5 km, whereas the 0-dBZ echo-top heights are relatively consistent throughout the day at about 6 km (Fig. 8b). The total cloud fraction from the ceilometer shows a persistent cloud fraction of about 0.3 throughout the day, consistent with the IC activity (Figs. 8a,c). Together these observations indicate overall limited vertical extent for the IC events and 0-dBZ echo tops consistent with that described by Ruppert and Johnson (2016).

On rainy days IC counts tend to be highest during the day whereas MCS counts tend to be highest at night, apart from the narrow peak in MCS counts near 1100 LT. The rain rates follow the pattern in MCS counts, being higher overall at night than during the day on rainy days, except for the late morning peak at 1000 LT lying between a local maximum in IC counts and MCS counts (Figs. 8a,b). Three days contributed to the large peak in rainfall at this time, two associated with Typhoons Cimaron and Jebi and the third associated with an MCS that passed very close to the ship on 28 September (Fig. 2a). These events were of high intensity, strongly affecting the mean rain rates at this hour. An analysis of the
representation of individual systems by the reanalyses is beyond the focus of this study.

The 30-dBZ echo tops have increased from an average of near 4–5 km on nonrainy days to 6 km on rainy days (Fig. 8b). The 0-dBZ echo tops are also overall higher on rainy days than on nonrainy days, reaching to 9 km at night and just below 8 km in the late afternoon. Total cloud fraction from the ceilometer is about double that seen on nonrainy days, but with maximum cloud cover at night and minimum during the day (Fig. 8c). The diurnal development of convection on rainy days seen in these panels is consistent with isolated convection in the late afternoon with generally lower vertical extent organizing into larger MCS systems at night and early morning, with the latter systems creating the greatest amount of rainfall, more extensive cloud cover, and deeper clouds.

The diurnal cycles in vertical velocity, rainfall, and cloud fraction for ERA5 and MERRA-2 conditioned on nonrainy and rainy days are shown in Fig. 9. The diurnal development of convection on rainy days seen in these panels is consistent with isolated convection in the late afternoon with generally lower vertical extent organizing into larger MCS systems at night and early morning, with the latter systems creating the greatest amount of rainfall, more extensive cloud cover, and deeper clouds.

The diurnal cycles in vertical velocity, rainfall, and cloud fraction for ERA5 and MERRA-2 conditioned on nonrainy and rainy days are shown in Fig. 9. The diurnal cycle in TGT ORG and SEA-POL rain rates respectively averaged from 5 to 120 km and from 5 to 10 km of the ship are also shown in all panels for reference, with SEA-POL rain rates scaled as in Fig. 8.

We have seen in Fig. 8 that even on days classified as nonrainy for the observations, isolated convection persists around the ship. Figure 9 indicates that more rain is present in the reanalyses on these days than in the observations, possibly reflecting the different spatial sampling of these products. However, reanalyses may also overestimate rainfall in the region (Fig. 2a). This is particularly indicated for ERA5, where rain rates have both an afternoon and early morning peak on nonrainy days coincident with peaks in grid-scale vertical velocity and cloud fraction suggestive of the growth of shallow afternoon convection into isolated deep convection with some stratiform clouds (Fig. 9a). A different picture is seen in MERRA-2, which shows rain steadily throughout the day, with only a weak diurnal enhancement at night (Fig. 9b). MERRA-2 indicates no low cloud development in the afternoon and only a small increase in middle-to-high cloud depth at night.

The lack of a diurnal cycle in low to middle cloud fraction in MERRA-2 is similar to that of the total cloud fraction at the ship on nonrainy days (Fig. 8c). While the vertical development of low clouds in the afternoon in ERA5 is in general agreement with the documented growth of daytime shallow cumulus clouds in the western Pacific under convectively suppressed conditions (Chen and Houze 1997; Johnson et al. 2001; Ruppert and Johnson 2015, 2016), this analysis suggests that the higher rainfall in ERA5 seen in Fig. 2a during PISTON is in part associated with more robust growth of daytime shallow convection on nonrainy days than seen in MERRA-2 or the observations.

Above the threshold for the observed cloud fractions, high clouds (above 7–8 km or ~450–400 hPa) are the most
Fig. 9. Reanalysis cloud fraction (shaded), omega (contour lines; Pa s$^{-1}$), and rain rate (pink line) for (a),(b) nonrainy and (c),(d) rainy days. Also shown in each panel are the TGT ORG and SEA-POL 120- and 10-km rain rates for reference.
abundant cloud type for both rain thresholds in the reanalyses (Fig. 9). On nonrainy days, with the exception of the peak high cloud around 14 km at 0200 LT likely associated with the growth of afternoon convection, the level of peak high cloud in ERA5 drops from about 14 km at midday to 10 km around 2000 LT, with maximum high cloud fraction around local noon and minimum high cloud fraction around midnight (Fig. 9a). The daytime maximum in high clouds seen in ERA5 is consistent with satellite and in situ observations, which suggest that shortwave heating during the day at upper levels associated with the clouds increases their depth and horizontal extent in the tropics (Chepfer et al. 2019; Ruppert and Klocke 2019; Wall et al. 2020). MERRA-2 high cloud fraction is more abundant and of greater vertical extent than that of ERA5 on nonrainy days, peaking at 12 km throughout the day with little diurnal variability in height or coverage (Fig. 9b). The overall higher positive relative humidity biases above 400 hPa (7 km) in MERRA-2 relative to ERA5 (Fig. 5c) may contribute to the greater fraction and depth of high cloud in this model on nonrainy days (Chepfer et al. 2019). High cloud biases above 400 hPa (7 km) may also contribute to the persistent weak upward motion at the base of the high clouds in MERRA-2, which is typically strongest when the depth of the high clouds is thickest, favoring upper-level overturning circulations and increased anvil cloud growth (Wall et al. 2020).

On rainy days, ERA5 rain rates, like the TGT ORG and SEA-POL 10-km rain rates, show a clear nighttime to early morning maximum, midday minimum, and steady increase from afternoon into the evening (Fig. 9c). MERRA-2 rainfall is sustained throughout the early morning hours and into the afternoon, with a minimum in the early evening in best overall agreement with the SEA-POL 120-km rain rates (Fig. 9d). As on nonrainy days, the amplitude of the MERRA-2 diurnal cycle in precipitation is significantly less than that of ERA5. The lower overall amplitude of the diurnal cycle may reflect the difference in horizontal resolution between ERA5 and MERRA-2, as a similar effect is seen when comparing the diurnal cycle in TGT ORG or SEA-POL 10-km rain rates with the diurnal cycle in SEA-POL 120-km rain rates.

Strong vertical motion (\( \leq -0.18 \text{ Pa s}^{-1} \)) is evident from 8 to 12 km throughout most of the diurnal cycle in ERA5 on rainy days, with high cloud fraction increasing downward in association with peaks in upward motion at these levels around 1000 and 2000 LT (Fig. 9c). Cloud fraction is a minimum at midlevels as seen on nonrainy days, while maximum low-level cloud depth and coverage is seen at night and into the early morning following the diurnal cycle in rainfall on these days. The diurnal maximum in upward motion from 8 to 11 km and increase in cloud cover at all levels in this model around 1000 LT corresponds to the observed maximum in the TGT ORG and SEA-POL rain rates for the three most intense events mentioned above, indicating possible impacts of data assimilation on the reanalyses. Outside of this midmorning convective activity, ERA5 shows strong upward motion (\( \leq -0.1 \text{ Pa s}^{-1} \)) below 3 km associated with the increase in cloud fraction starting around 2000 LT, with the upward motion elevated to above 5 km after midnight, suggestive of latent heating from the growth of early morning stratiform convection in the model.

Like ERA5, low-level cloud fraction and depth are maximum at night in MERRA-2 consistent with peak rainfall in this model, although peak values are less than those for ERA5. MERRA-2 also shows an increase in midlevel cloud fraction from 4 to 6 km and a peak in upward motion just after midnight, consistent with the development of deep convection in the model. As in Fig. 7b, MERRA-2 lacks the strong top-heavy heating seen in ERA5 on rainy days. The afternoon peak in upward motion at the base of the high cloud layer in MERRA-2 is similar to that seen in ERA5 on nonrainy days and is likely related to anvil cloud circulations in MERRA-2. In addition, a minimum in mid- and low-level clouds and upward motion below 6 km is seen at this time, suggesting that the increased high cloud fraction is not associated with afternoon deep convection in the model.

b. Surface forcing

The diurnal cycles in surface heat and radiative fluxes, rainfall, surface to 10-m moisture (\( \Delta q \)) and temperature (\( \Delta T \)) gradients, and 10-m wind speed for nonrainy and rainy days are shown in Fig. 10. As in Fig. 2b, the TGT fluxes calculated using the COARE bulk formula are used for comparison with ERA5 and MERRA-2 surface fluxes provided by the modeling centers. The surface to 10-m moisture gradient is calculated as the difference between the saturated mixing ratio based on SST and surface pressure and the 10-m mixing ratio for both TGT and the reanalyses. Similarly, the surface to 10-m temperature gradient is calculated as the difference between SST and 10-m air temperature. A positive gradient in moisture or temperature is associated with a more negative (greater cooling) heat flux.

On nonrainy days the observed SHF indicates a slight maximum cooling in the afternoon coincident with higher daytime 10-m wind speed and a slight maximum in \( \Delta T \) at this time (Figs. 10a,c). The observed LHF is considerably larger than the SHF but follows a similar pattern on nonrainy days (Fig. 10b), with maximum values in the afternoon when wind speeds are maximum. A small diurnal maximum in \( \Delta q \) is also present (Fig. 10c). A small decrease in SWD (Fig. 10d) and increase in LWD (Fig. 10e) in the afternoon on nonrainy days, coincident with an increase in cloud fraction (Fig. 8e), suggests that nonprecipitating clouds form around 1500 LT, when the surface moisture flux is at its diurnal peak. This relationship between surface fluxes and the diurnal cycle in cumulus clouds agrees with previous studies of the western Pacific warm pool under convectively suppressed conditions (Ruppert and Johnson 2016).

Overall, both reanalyses overestimate SHF on nonrainy days relative to the observations. As noted in section 3b, the SHF bias in MERRA-2 is improved using the COARE bulk flux algorithm while this is not the case for ERA5 (not shown). Figure 10a shows that diurnal peaks in SHF for ERA5 in the early morning and midday are coincident with periods of convection in this model (Fig. 9a). In addition, biases in 10-m air temperature seen in Fig. 5f result in larger
near-surface temperature gradients than what is observed (Fig. 10c). These factors may explain why ERA5 SHF biases are less sensitive to the choice of bulk formula when compared with MERRA-2. Both reanalyses also overestimate LHF on nonrainy days but capture the diurnal cycle in this parameter reasonably well (Fig. 10b). As for SHF, the overestimation of LHF in MERRA-2 relative to the observations is greatly reduced by using the COARE bulk flux algorithm. For ERA5 the reasons for the observed bias are less apparent and, as for SHF, may be the result of more convective activity in ERA5 on nonrainy days relative to the observations. On the other hand, it is possible that the high SHF and LHF in ERA5 contribute to cloud development in this model on these days. While MERRA-2 also overestimates these fluxes, cloud vertical growth could be suppressed because of the subsidence and dry air above the boundary layer in MERRA-2 on nonrainy days (Figs. 5d and 9b).

The abundance of high clouds in MERRA-2, particularly from sunrise to late afternoon (Fig. 9b), results in lower SWD on nonrainy days relative to the observations (Fig. 10d). ERA5 also underestimates the SWD at the surface, but in the case of this model the SWD bias likely results from precipitating clouds in ERA5 on nonrainy days. Low clouds in MERRA-2 have low coverage and are thin throughout most of the diurnal cycle on nonrainy days (Fig. 5d), trapping less infrared radiation and resulting in lower LWD than what is observed (Fig. 10e). In contrast, low clouds in ERA5 are deeper than those in MERRA-2, becoming precipitating clouds in the afternoon and early morning, and resulting in higher LWD than in the observations, particularly at night.

On rainy days the observed SHF and LHF increase primarily as a result of increased wind speed, although increased air–sea temperature differences also act to increase SHFs on these days (Figs. 10f–h). As wind speed remains consistent throughout the day on rainy days, LHF tends to have a smaller diurnal cycle with a maximum during the day and minimum at night, following the modest diurnal cycle in $D_q$. The observed SHF has a larger-amplitude diurnal cycle on rainy days than on nonrainy days, being maximum from early morning to late afternoon and minimum in the early evening, following the diurnal cycle in air–sea temperature differences (primarily associated with 10-m air temperature) (Fig. 10h) and convective activity (Fig. 10i). Enhanced cloudiness throughout the day reduces SWD and increases LWD in

---

**Fig. 10.** Diurnal cycle in (a),(f) SHF; (b),(g) LHF; (c),(h) SST $- T_{10m}$ (dashed thin line), $q_s - q_{10m}$ (dotted thin line), and $W_{10m}$ (solid thin line); (d),(i) SWD (thick line) and rain rate (thin line); and (e),(j) LWD for (left) nonrainy and (right) rainy days. The TGT rain rates in (d) and (i) are from ORG. Error bars are standard deviation of the mean and are shown every 3 h. Only error bars for TGT are shown in (c) and (h). Note the difference in scale for wind speed (right y axis) in (c) and (h).
the observations relative to nonrainy days (Figs. 10i,j). The latter suggests low cloud fraction has increased over nonrainy days at the ship, with lowest cloud bases/optically thickest clouds at 0400 and 1200–1600 LT when MCS counts and IC counts, respectively, are at their overall diurnal maximum (Figs. 8a,b), and highest cloud bases/optically thinnest clouds around 0800 LT when IC cloud counts are maximum and MCS counts are minimum (Fig. 8a,b). The observed surface forcing suggests LHFs continuously replenish moisture in the boundary layer on active days, while SHFs draw heat out of the ocean primarily during peak convective periods.

Despite similar 10-m wind speeds, the negative 10-m air temperature bias in ERA5 persists on rainy days resulting in a larger near-surface temperature gradient relative to what is observed (Fig. 10h) and similar biases in SHF as seen on nonrainy days, while MERRA-2 SHFs are in better agreement with the observations on these days (Fig. 10f). Both reanalyses show better overall agreement with the observed LHF on rainy days, with biases reduced relative to nonrainy days throughout most of the day (Fig. 10g). The radiative fluxes on rainy days indicate that while both the observations and reanalyses show lower SWD and higher LWD consistent with more cloud cover with lower cloud bases, MERRA-2 LWD is significantly less than ERA5 and TGT, consistent with MERRA-2 cloud bases remaining higher than ERA5 or the observations. Figure 9d indicates that both optically thinner clouds and those with higher cloud bases are in fact present in MERRA-2 when compared with ERA5 on these days. As with nonrainy days, biases in moisture above the boundary layer in MERRA-2 seen in Figs. 5d and 5e likely contribute to the lack of low cloud development in this model.

6. Summary and conclusions

This study evaluates ERA5 and MERRA-2 rainfall, cloudiness, and associated parameters using high-quality in situ observations collected during the 2018 PISTON field campaign. The northern equatorial western Pacific was dominated by synoptic-scale wave and typhoon activity during PISTON, with weak monsoon, MJO, and other convectively coupled equatorial wave activity. Reanalyses are widely used for improving our understanding of tropical weather and climate, and tropical convection is an important aspect of these studies as it drives large-scale vertical motions through latent and radiative heating of the column. The main questions addressed by this study are the following: 1) How well do the gridded products represent the background state of the northern equatorial western Pacific during the 2018 PISTON campaign? 2) How well do these products capture the diurnal cycle of clouds and rainfall in this region?

Comparison with the roughly 3-hourly radiosonde launches from the TGT during PISTON reveals biases in thermodynamic quantities at critical levels in the atmosphere. ERA5 and MERRA-2 have similar structure to their moisture biases with height, with negative moisture biases in the lower troposphere and positive biases to 200 hPa. The vertical structure in temperature biases is less similar, with ERA5 having a maximum negative temperature bias of about −1°C near the surface and at 100 hPa, while biases of less than one to two tenths of a degree are seen elsewhere in the profile. In contrast, MERRA-2 negative temperature biases are larger than a few tenths of a degree throughout the profile, with the largest negative biases also on the order of 1°C seen at the top of the ML and in the midtroposphere, and a similar magnitude positive bias seen in the upper troposphere. The smaller moisture and temperature biases in ERA5 than in MERRA-2 may in part be the result of ERA5 having assimilated PISTON observations. A data-denial experiment in which ERA5 is processed without PISTON data would be needed to confirm this hypothesis.

Biases in temperature and moisture are associated with biases in high and low cloud fractions in the reanalyses, particularly through their impact on relative humidity. Cloud biases can then feed back on the temperature and moisture profiles as noted by Wright et al. (2020). At upper levels, high cloud fractions (above 7 km, or 400 hPa) are higher on average in ERA5 than MERRA-2 (Fig. 4d) but differ both in their rainy versus nonrainy coverage (Fig. 7), as well as across the diurnal cycle (Fig. 9). The reanalysis daytime SWD is less than that at the ship on both nonrainy and rainy days, suggesting that the reanalyses may overestimate high cloud regardless of the convective conditions. These biases are hypothesized to be the result of the large positive biases in relative humidity at upper levels, which, together with the model critical RH value, determine large-scale cloud fraction (e.g., Quaas 2012; Miao et al. 2019; Wright et al. 2020). Biases in MERRA-2 high cloud and upper-level moisture are also reported elsewhere (Miao et al. 2019; Wright et al. 2020). At mid- and low levels, cold, dry biases above the mixed layer (Figs. 5d–f) result in greater low-level stability (Fig. 6c). The larger dry bias in MERRA-2 above the boundary layer relative to ERA5 results in fewer and thinner low clouds with higher cloud base and continuous light rain on nonrainy days, with heavier rain seen particularly at night on rainy days. ERA5 low clouds show more vertical development to midlevels than MERRA-2 and are generally associated with higher rain rates and larger diurnal variability on both nonrainy and rainy days.

Weak upward motion is seen in ERA5 during nonrainy periods, while weak downward motion is seen in MERRA-2 throughout most of the troposphere, suggesting more suppressed conditions exist away from precipitating systems in MERRA-2 in the PISTON region. Composites of vertical velocity under rainy conditions indicate that ERA5 captures the expected top-heavy latent heating associated with periods of observed MCSs near the ship and that these systems have more intense heating translating to stronger grid-scale vertical motions than those of the IC composite. In contrast, MERRA-2 vertical velocity profiles have a similar vertical structure for both the MCS and IC composites, suggesting that MERRA-2 did not produce the expected strong stratiform heating expected for the observed periods of MCS activity. Analysis of the diurnal cycle in vertical velocity for the models (Fig. 9) indicates that peak upward motion at upper levels during the day is in part associated with radiatively driven circulations at the base of model high clouds, particularly for MERRA-2. The maximum vertical
velocity associated with the nighttime convection in MERRA-2 is limited to midlevels and is less than half the magnitude of that seen in ERA5, further suggesting convective heating peaks at lower levels in MERRA-2 than ERA5. This result has implications for studies that rely on reanalysis to examine the relationship between heating profiles in the tropics and the large-scale circulation (e.g., Schumacher et al. 2004).

The mean biases in tropopause specific humidity and temperature in the reanalyses observed during PISTON are consistent with the documented biases for these products reported elsewhere (Gelaro et al. 2017; Hersbach et al. 2020; Wolding et al. 2022). In particular, Hersbach et al. (2020) find that global moisture increments in ERA5 are on the order of 1%–2% or 1–1.5 g kg\(^{-1}\) near 850 hPa, while those for MERRA-2 in the global tropics are about double this amount (Gelaro et al. 2017). Similarly, global temperature increments in ERA5 have maximum amplitudes of about 0.1 K in the troposphere (Hersbach et al. 2020), whereas those for MERRA-2 are as high as 0.4–0.5 K (Gelaro et al. 2017). The current study finds larger biases in both moisture and temperature for MERRA-2 in the PISTON region (Figs. 5d,f), consistent with the differences in the magnitude of the temperature and moisture increments found in these two studies. The assimilation of microwave imager data after 1998, which has been shown to cause warming and drying in models at 850 hPa (Hersbach et al. 2020), may have contributed to the dry bias in both ERA5 and MERRA-2. Errors in the in situ radiosonde observations must also be considered (e.g., Wolding et al. 2022). However, as noted by Wolding et al. (2022) and references therein, even considering observational and sampling errors, the biases in moisture and temperature in the reanalyses suggest that different processes are occurring in the lower troposphere than those of the observations. The results of the present study, like those of Wolding et al. (2022), support the National Academy of Science’s recent report on decadal strategies for observations from space (NAS 2018), which emphasizes the need for observations in the lower troposphere, particularly over open ocean regions, to constrain models and improve their reliability in process-oriented studies.

Acknowledgments. We thank Paul Ciesielski, Colorado State University, for kindly providing the quality-controlled radiosonde data for this study. We also thank Elizabeth Thompson, from NOAA PSL, for providing the quality-controlled PISTON surface meteorological, ceilometer, and air–sea flux data used for this study. Thanks also are given to all of the personnel who participated in PISTON and made these data available to the broader community, as well as to two anonymous reviewers whose comments greatly improved the quality of this paper. Authors Yolande Serra, Steven Rutledge, and Kyle Chudler were supported by the Office of Naval Research Award 12076040. This is PMEL Contribution 5340. This publication is partially funded by the Cooperative Institute for Climate, Ocean, and Ecosystem Studies (CIOCES) under NOAA Cooperative Agreement NA20OAR4320271, Contribution 2022-1195.

Data availability statement. The SEA-POL and radiosonde data from CSU used for this study are available online (https://doi.org/10.5067/SUBORBITAL/PISTON2018-ONR-NOAA/RVT THOMPSON/DATA001). NOAA PSL surface meteorological and ceilometer data were not available at this DOI at the time of publication of this paper but will be released upon completion of quality control. To obtain these data before their official release contact Elizabeth Thompson at Elizabeth.Thompson@noaa.gov. The Goddard Earth Sciences Data and Information Services Center (GES DISC), provided the TRMM 3B42 (Huffman et al. 2016) and MERRA-2 (GMAO 2008a,b, 2015a,b) datasets for this study. The Hersbach et al. (2018a,b) data were downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store. The results contain modified Copernicus Climate Change Service information. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or the data that it contains.

REFERENCES


——, ——, J. E. Hare, A. A. Grachev, and J. B. Edson, 2003: Bulk parameterization of air–sea fluxes: Updates and


GMAO, 2008b: tavg1_2d_slv_Nx: MERRA 2D IAU diagnostic, single level meteorology, time average 1-hourly V5.2.0. GES DISC, accessed 18 November 2019, https://doi.org/10.5067/B6DQZQLSFDLH.


