Special Collection
Process Diagnostics for CMIP6

Process-Oriented Diagnosis of Tropical Cyclones in Reanalyses Using a Moist Static Energy Variance Budget

CAITLIN A. DIRKES,[a] ALLISON A. WING,[a] SUZANA J. CAMARGO,[b] AND DAEHYUN KIM[c]

[a] *Department of Earth, Ocean and Atmospheric Science, Florida State University, Tallahassee, Florida*
[b] *Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York*
[c] *Department of Atmospheric Sciences, University of Washington, Seattle, Washington*

(Manuscript received 20 May 2022, in final form 6 April 2023, accepted 7 April 2023)

ABSTRACT: Global models are frequently used for tropical cyclone (TC) prediction and climate projections but have biases in their representation of TCs that are not fully understood. The objective of this work is to assess how well and how robustly physical processes that are important for TC development are represented in modern reanalysis products and to consider whether reanalyses can serve as an observationally constrained reference against which model representation of these physical processes can be evaluated. Differences in the representation of large-scale environmental variables relevant to TC development do not readily explain the spread in TC climatologies across climate models, as found in prior work, or across reanalysis datasets, as shown here. This motivates the use of process-oriented diagnostics that focus on how convection, moisture, clouds, and related processes are coupled and can be used to identify areas to target for model improvement. Using the column-integrated moist static energy (MSE) variance budget, we analyze radiative and surface flux feedbacks across five different reanalyses. We construct an intensity-bin composite of the MSE variance budget to compare storms of similar intensity. Our results point to some fundamental differences across reanalyses in how they represent MSE variance and surface flux and radiative feedbacks in TCs, which could contribute to differences across reanalyses in how they represent TCs, but other factors also likely contribute. Any future work that evaluates these diagnostics in GCMs against reanalyses should do so cautiously, and efforts should be undertaken to provide a true observational estimate of these processes.

KEYWORDS: Tropical cyclones; Reanalysis data; Heat budgets/fluxes; Moisture/moisture budget

1. Introduction

Tropical cyclones (TCs) have significant social, economic, and infrastructural impacts on communities around the world each year, heightening the importance of their accurate representation in global models. Simulation of TCs in general circulation models (GCMs) has made tremendous strides in the last few decades, with improved performance in GCMs (Camargo and Wing 2016). However, there are still biases in these models (Walsh et al. 2016; Roberts et al. 2020). For example, there are often too few storms simulated in the Atlantic and too many in the Indian Ocean (Camargo 2013; Roberts et al. 2020). Simulated TCs are also typically too weak compared to real-world storms (Shaevitz et al. 2014). Prior research has shown that higher resolution improves the fidelity of the simulated TC climatology, including the number of TCs and their structure, rainfall, and intensity (W. Zhang et al. 2021; Vidale et al. 2021; Roberts et al. 2020; Manganello et al. 2012; Wehner et al. 2014; Murakami et al. 2015). For example, Moon et al. (2020b) found that GCMs with finer horizontal resolutions simulate more realistic radii of maximum winds and, therefore, inner-core structures. However, even across GCMs at the same resolution, fundamental differences remain in the way TCs are simulated (Roberts et al. 2015; Shaevitz et al. 2014; Kim et al. 2018; Moon et al. 2020b; Wing et al. 2019; Russotto et al. 2022).

TCs evolve via interactions with their environment and are tightly coupled to convective processes; thus, the representation of these processes in models influences TC simulation. In particular, TC simulation is sensitive to the model dynamical core and convective and cloud microphysics parameterizations (Reed and Jablonowski 2011; Murakami et al. 2012; Zhao et al. 2012; Kim et al. 2012; Reed et al. 2015; Duvel et al. 2017; Russotto et al. 2022). Variation in the large-scale environment, as captured by genesis potential indices (Emanuel and Nolan 2004; Emanuel 2010; Tippett et al. 2011), has helped explain TC variability on intraseasonal, interannual,

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-22-0384.s1.

Corresponding author: Allison A. Wing, awing@fsu.edu

DOI: 10.1175/JCLI-D-22-0384.1

and climate change time scales (Camargo et al. 2007a,b, 2014, 2016). However, Camargo et al. (2020) showed that the intermodel spread in the historical TC climatology across models is not explained by differences in the simulated large-scale environment. Therefore, other processes must determine the climatology of simulated TCs in GCMs of a given resolution.

Process-oriented diagnostics (PODs) are used to characterize a specific physical process, often associated with model physics and numerics, that is related to the ability to simulate an observed phenomenon, such as a TC (Maloney et al. 2019). This approach goes beyond simply evaluating the number and intensity of TCs and diagnoses the underlying mechanisms relevant to the storm development process. PODs inform which areas to target for model improvement and can be compared to observations, revealing whether simulated TC development occurs for the right physical reasons. Kim et al. (2018) was the first study to develop a POD for TCs in climate models. They used a spatial composite approach to isolate TC kinematic and thermodynamic structures during intensification, focusing on how convection, moisture, clouds, and related processes are coupled at the grid scale. Wing et al. (2019) further quantified these processes by adapting the column-integrated moist static energy (MSE) variance budget, originally developed for studies of convective self-aggregation, for GCM simulation of TCs. Application of these diagnostics to historical GCM simulations revealed that surface flux feedbacks are responsible for part of the intermodel spread in TC intensity and that GCMs that produce more intense TCs also produce a greater amount of precipitation in the TC inner core (Kim et al. 2018; Wing et al. 2019; Moon et al. 2020b). Wing et al. (2019) and B. Zhang et al. (2021) also found that radiative feedbacks in GCMs are relatively more important to TC development in weak storms. However, in comparison to satellite observations, most GCMs overestimate the amount of inner-core precipitation. This suggests that in order to simulate TCs of a certain intensity, GCMs need to develop large rain rates (and thus diabatic heating) in the inner core, perhaps to compensate for deficiencies in their representation of other processes (Vannière et al. 2020; Moon et al. 2022). Application of these diagnostics may aid in model development by identifying processes relevant for TCs that may be misrepresented; for example, if the surface flux feedback is diagnosed to be too weak, a model developer might reexamine the dependence of surface fluxes on wind speed in their boundary layer scheme.

The primary objective of this paper is to extend the work of Wing et al. (2019) to assess how some processes that are important for TC development are represented in modern reanalysis products and to consider whether reanalysis can serve as an observationally constrained reference for the MSE variance budget diagnostic for TCs against which GCMs could be evaluated.

Reanalyses employ numerical weather prediction models with assimilated observations to provide a self-consistent estimate of the atmospheric state without temporal and spatial gaps (Parker 2016). Reanalyses are only partially constrained by observations but are the only dataset that is able to provide all the necessary variables, with sufficient temporal and spatial resolution, to perform a full calculation of the MSE variance budget around TCs. However, it is important to note that reanalysis representation of TCs is, itself, subject to errors and biases. While reanalyses can reproduce most of the features of the observed TC climatology, as well as realistic variability on interannual and intraseasonal time scales, they typically underestimate TC intensity and exhibit TC position errors compared to observed best-track data (Schenkel and Hart 2012; Murakami 2014; Hodges et al. 2017; Kim et al. 2021). Reanalyses also misrepresent TC outer size and structure, the relationship between intensity and outer size, and the TC life cycle (Schenkel and Hart 2012; Schenkel et al. 2017; Bian et al. 2021), as well as differ from observations in their representation of physical processes such as the precipitation generated by TCs (Jones et al. 2021). These discrepancies may be caused by coarse horizontal grid spacing, parameterization of subgrid-scale processes, data assimilation schemes, vortex relocation or initialization techniques utilized by each dataset, or other model details. Regardless of resolution, reanalyses struggle to find a middle ground between simulating enough storms compared to observations and simulating them at the proper intensity, which is likely associated with the individual models' physics (Aarons et al. 2021). Reanalyses also may either miss storms that happened in the real world (especially weak storms) or simulate artificial storms that did not happen in the real world (Hodges et al. 2003; Zarzycki et al. 2021). These biases motivate our use of five different modern reanalysis datasets, as no single dataset represents truth, and we assess how well the reanalyses agree on the representation of the TC MSE variance budget to contextualize any future GCM-reanalysis comparisons.

A secondary objective of this work is to investigate whether differences in the representation of TCs across reanalyses can be explained by the storm-scale physical processes captured by the MSE variance budget or by the climatology of the large-scale environment. While the MSE variance budget was developed as a process-oriented diagnostic for climate models (Wing et al. 2019), it is plausible to consider whether it also explains differences in TC representation across reanalyses. In addition, while large-scale environmental factors do not explain differences in TCs across GCMs (Camargo et al. 2020), it has not yet been shown whether this is true for reanalyses as well. Other factors, such as how synoptic-scale environments impact storm-level processes (McBride and Zehr 1981; Elsberry et al. 1988; Jones 1995; DeMaria 1996; Hill and Lackmann 2009), may also influence reanalysis representation of TCs (Slocum et al. 2022) but are not examined here.

The remainder of this paper is structured as follows: section 2 describes the datasets used and analysis methods. To provide context and explore a possible reason for spread in the TC climatology across reanalyses, section 3 compares aspects of the climatological large-scale environment relevant to TCs. Section 4 presents the MSE variance budget diagnostics, their spread across reanalyses, and their relationship to aspects of the TC climatology. The results and the implications for their use as a process-oriented diagnostic are summarized in section 5. Several sensitivity tests are presented in appendixes A and B.

TABLE 1. Summary of the reanalysis datasets and their model characteristics, including resolution (latitude × longitude), data assimilation scheme (DA), any TC preprocessing that is performed prior to data assimilation, and relevant parameterizations. Model resolution is shown based on the model spectral resolution, with the approximate horizontal resolution in parentheses (all use spectral dynamical cores except for MERRA-2, which uses a finite-volume dynamical core). Note that CFSR is a coupled atmosphere–ocean–sea ice reanalysis system.

Reanalysis	Model	Model resolution	Output resolution	DA	Preprocessing
MERRA-2	GEOS 5.12.4 (2015)	0.50° × 0.625° 72 levels	0.50° × 0.625°	3DVar IAU	Vortex relocation
CFSR	NCEP CFS (2007)	T382 (~0.34°) 64 levels	0.50° × 0.50°	3DVar	Vortex relocation
JRA-55	JMA GSM (2009)	T319 (~0.50°) 60 levels	0.56° × 0.56°	4DVar	TC wind-profile retrieval
ERA5	IFS Cycle 41r2 (2016)	T639 (~0.28°) 137 levels	0.25° × 0.25°	4DVar	None
ERA-Interim	IFS Cycle 31r2 (2007)	T255 (~0.54°) 60 levels	0.7018° × 0.7031°	4DVar	None

Reanalysis	Convective parameterization	Cloud scheme	Radiation scheme
MERRA-2	Moorthi and Suarez (1992)	Bacmeister et al. (2006)	Chou and Suarez (1999) Chou et al. (2001)
CFSR	Tiedtke (1983) Moorthi et al. (2001)	Xu and Randall (1996) Zhao and Carr (1997)	Clough et al. (2005)
JRA-55	Arakawa and Schubert (1974) Xie and Zhang (2000)	Kawai and Inoue (2006)	Briegleb (1992); Chou et al. (2001) Freidenreich and Ramaswamy (1999)
ERA5	Tiedtke (1989)	Tiedtke (1993)	Mlawer et al. (1997)
ERA-Interim	Tiedtke (1989)	Tiedtke (1993)	Fouquart and Bonnel (1980) Mlawer et al. (1997)

2. Data and methods

a. Data

1) REANALYSIS DATASETS

Five reanalysis datasets are used in this study: the National Aeronautics and Space Administration (NASA) Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2; Gelaro et al. 2017), the National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR; Saha et al. 2010c, 2014), the Japanese 55-year Reanalysis (JRA-55; Kobayashi et al. 2015), the fifth major global reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) (ERA5; Hersbach et al. 2020), and the ECMWF interim reanalysis (ERA-Interim; Dee et al. 2011). Table 1 describes the characteristics of each dataset; Fujiwara et al. (2017) also provides a useful review. ERA-Interim output is provided on a Gaussian grid, but since the latitude and longitude spacings are constant within our analysis region in the tropics, we treat it as a regular grid with 0.7018° × 0.7031° spacing. Our analysis considers 1980–2016. For analysis of the large-scale environment [sections 2b(1) and 3], we interpolate all fields to a common 2° grid for comparison. For the MSE variance budget [sections 2b(2) and 4], we consider each reanalysis on its own grid. The native model resolution JRA-55 data are provided on native hybrid sigma–pressure model levels, so we first interpolate it to pressure levels; for all other reanalyses, we use the data provided on pressure levels. We test the sensitivity of the MSE variance budget in JRA-55 to using 1.25° × 1.25° output data rather than the native model resolution data (gridded to 0.56° × 0.56°) in appendix A.

In general, reanalyses use one of two data assimilation (DA) schemes: three-dimensional variational assimilation (3DVar) and four-dimensional variational assimilation (4DVar). 3DVar assimilates observations by assuming that all observations used within the analysis period were recorded at the initialization time of the model and, thus, does not account for temporal variability other than through sequential assimilation of observations. On the other hand, 4DVar does account for different observation times and allows for a flow-dependent influence of observations to constrain the model (Thépaut et al. 1996; Dee et al. 2011). This leads to lower error propagation during forecast generation than does 3DVar, and so is generally the data assimilation scheme used by more recent reanalysis products. We also note that MERRA-2 utilizes an incremental analysis update procedure (IAU), in which the analysis increments are applied gradually as a correction to the background state at each time step rather than abruptly at the analysis time (Gelaro et al. 2017; Bloom et al. 1996). This may provide better consistency

among the different fields. However, we utilize MERRA-2's analysis (ANA) products, rather than the assimilation (ASM) products that are the result of applying the IAU procedure, because the former are more consistent with the analyses produced by other reanalysis systems. Three of the reanalyses feature some sort of preprocessing method to aid in TC tracking and representation efforts. Both CFSR (Saha et al. 2010c) and MERRA-2 (McCarty et al. 2016) have vortex relocation, which is when TC vortices are moved from their initial reanalysis-determined location to the best track position provided by the National Hurricane Center (NHC) and Joint Typhoon Warning Center (JTWC) prior to data assimilation of storm observations. JRA-55 is unique in that it generates synthetic TC wind profiles from historical observations and assimilates them into the reanalysis as if they are dropsonde observations for TCs with 10-m wind speeds greater than 34 kt (Kobayashi et al. 2015), where 1 kt $\approx$ 0.514 m s^{-1}. Both of these methods of TC preprocessing likely aid in the accuracy of TC simulation (Schenkel and Hart 2012). In addition to those discussed here, there are other differences between reanalyses, such as the choice of prognostic variables, form of model equations, and numerical methods. For example, while ERA5 and ERA-Interim use the same modeling system, one of several differences between them is a change in the prognostic equations for moisture variables. While we have noted some of the important characteristics of the different reanalyses here for context, as discussed in section 5, we do not attempt to attribute our results to particular components of the reanalysis configurations, since the sheer number of differences makes it impossible to isolate each of their influences without doing targeted sensitivity experiments with individual models.

What we refer to as CFSR is actually a combination of CFSR (Saha et al. 2010c) from 1979 to 2010 and then CFSv2 (Saha et al. 2014) from 2011 to 2016, where the latter provides data using the same model that extends the CFSR reanalysis beyond December 2010. There are a few select dates during our analysis period for which CFSR is missing data and are thus ignored: 15 May 2013, 16 June 2013, 17 June 2013, 29 March 2016, 31 March 2016, 1 April 2016, 2 April 2016, and 17 December 2016. In addition, CFSR is known to exhibit discontinuities in the mass field, which may lead to potential imbalance with the wind field (Schenkel and Hart 2012). Other reanalyses have other known data issues: there some instances of nonphysical 10-m wind speeds exceeding 100 m s^{-1} in ERA5 (Hersbach et al. 2020) and a few TCs in JRA-55 that are misrepresented as anticyclonic vortices before 1987, mostly in the North Atlantic (JMA 2020). These instances are not included in our analyses. Finally, there are also a few isolated instances for which an ERA5 variable used in our calculations was corrupted: specific humidity at 1200 UTC 2 May 1980, and sensible heat flux at 0600 UTC 8 December 1990 and 1200 UTC 10 July 1999. We became aware of these corrupt times after the majority of our analysis had been completed; therefore, we left them incorporated in the analysis. These three corrupt data points are not expected to affect the results, due to the large sample size over which we take composites, as discussed later in this section.

2) TC TRACKS

We use TC tracks from Zarzycki et al. (2021), which are derived from an objective tracking algorithm, TempestExtremes, applied to reanalyses. TempestExtremes identifies a TC center by finding a mean sea level pressure that is surrounded within 5° by a closed contour that is 2 hPa higher. Then, it eliminates the tracking of cold-core cyclones by requiring a 300–500-hPa geopotential thickness maximum to be located horizontally within 1° of the identified TC center (Ullrich and Zarzycki 2017; Zarzycki et al. 2017). According to Zarzycki et al. (2021), JRA-55 has the best statistics for representing TCs that also exist in the best track data. Conversely, MERRA-2 has the highest false alarm rate, which means it tracked the most TCs with no official record in the postseason analysis. The representation of TCs in MERRA-2 is discussed in more detail by Aarons et al. (2021).

Figure 1 shows the TCs tracked by TempestExtremes in the reanalysis datasets; the tracks in Fig. 1f are from the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al. 2010), which is considered the "actual" TC track data from a postseason analysis. As noted by Kim et al. (2021), MERRA-2 has a notable equatorward bias in its TC track distribution, especially in the eastern Pacific and Southern Hemisphere. The reanalyses capture the general spatial distribution of TCs but do not include every TC that happens in reality, nor do they simulate them at the proper intensity (Fig. 2; Murakami 2014; Hodges et al. 2017; Zarzycki et al. 2017; Aarons et al. 2021). This underperformance is a general limitation of reanalyses that may imprint on our results. Of the reanalyses, CFSR, MERRA-2, and JRA-55 simulate the strongest storms, while ERA-Interim and ERA5 simulate weaker storms, but none of the reanalyses capture the most intense storms (>60 m s^{-1}) seen in observations. It is notable that the reanalyses that simulate the strongest storms all have TC preprocessing methods. ERA-Interim and ERA5 have similar intensity probability density functions despite the much higher resolution of ERA5. Since both are derived from versions of the ECWMF model with similar subgrid-scale parameterizations, we might have expected that the higher resolution in ERA5 would have resulted in a higher proportion of stronger storms, though even 0.25° resolution is not high enough to expect to simulate category 4 and 5 storms on the Saffir–Simpson hurricane wind scale (Moon et al. 2020a; Davis 2018; Walsh et al. 2007).

b. *Methodology*

1) CLIMATOLOGICAL LARGE-SCALE ENVIRONMENT

We consider several TC-related, climatological, large-scale environmental variables as possible sources of inter-reanalyses spread in TC representation. All environmental fields are linearly interpolated to a common 2° grid for comparison and are computed from monthly mean fields as climatologies from 1980 to 2016 averaged over the peak TC season: August–October in the Northern Hemisphere and January–March in the Southern Hemisphere. When we consider spatial averages of the environmental fields, we average over ocean regions equatorward of

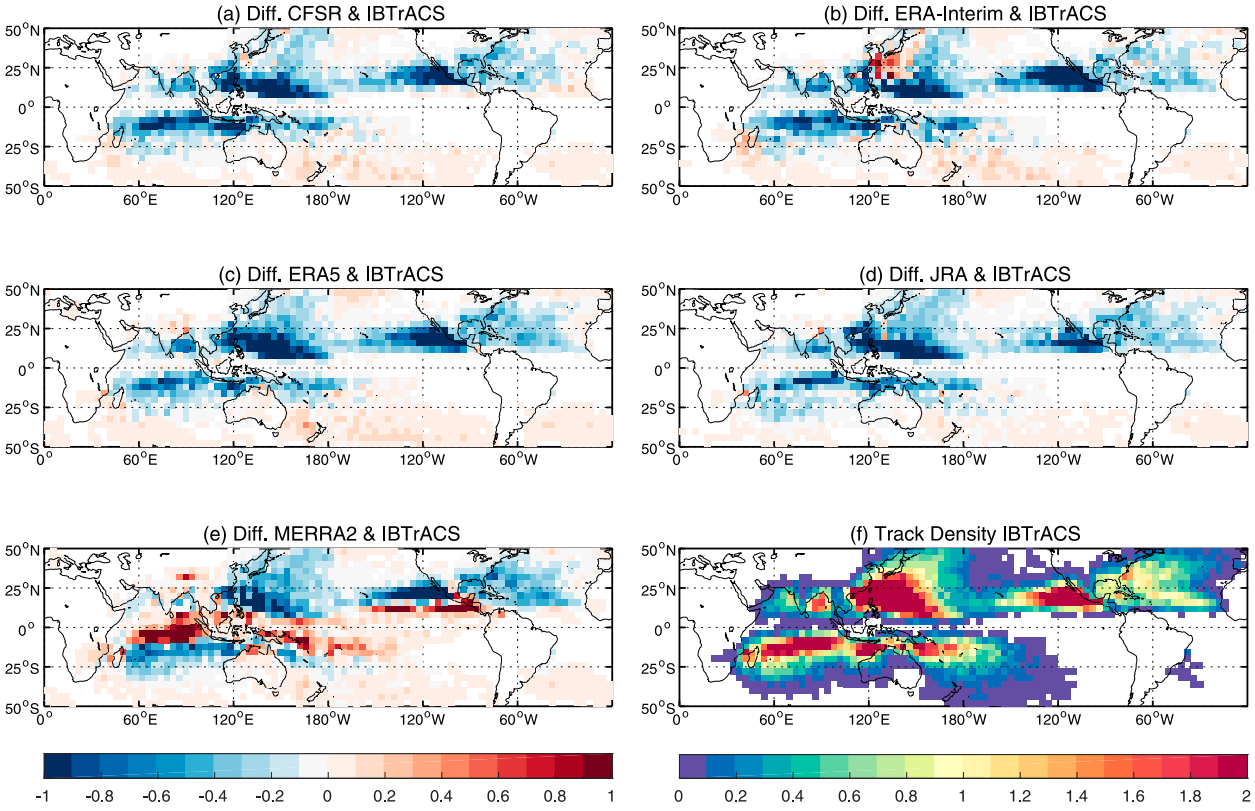

FIG. 1. (a)–(e) Difference between the track density climatology per year from TempestExtremes for the reanalysis datasets and IBTrACS. (f) IBTrACS track density (tropical storm intensity and above) for the period 1980–2016.

30°, excluding the South Atlantic and eastern South Pacific, where there is minimal observed TC activity. As noted above, genesis potential indices (Emanuel and Nolan 2004; Emanuel 2010; Tippett et al. 2011; Wang and Murakami 2020) capture aspects of the large-scale environment that are important for TCs.

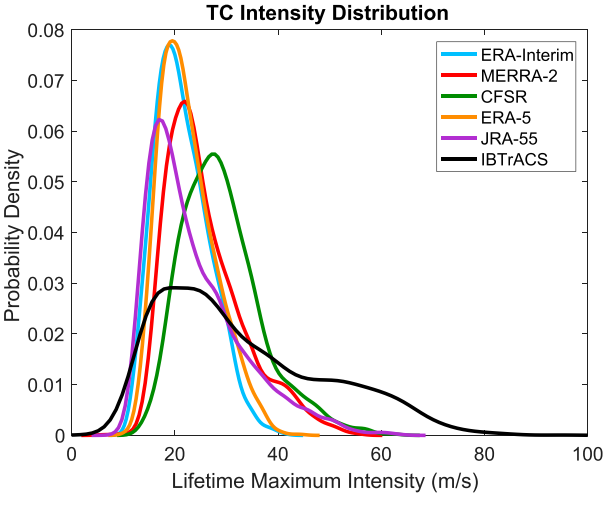

FIG. 2. Distribution of lifetime maximum wind speed for TCs included in our composites in reanalyses and observations (IBTrACS; Knapp et al. 2010) from 1980 to 2016.

We use the TC genesis index (TCGI), which is an empirical genesis index following the formulation of Tippett et al. (2011) and Camargo et al. (2014), in which the integrated value of TCGI gives the predicted number of TCs by the index. It is based on a Poisson regression fit for each of the reanalyses using their climatological fields and the TC historical climatology for storms whose lifetime maximum intensity is greater than or equal to 35 kt. In the main text, we show a version of TCGI that uses column relative humidity as its humidity variable, but we also consider a version of TCGI that uses saturation deficit as its humidity variable as well as an alternate index, the genesis potential index (GPI; Emanuel and Nolan 2004; Camargo et al. 2007a) (Fig. S1 in the online supplemental material). We also examine other individual thermodynamic and dynamic environmental variables, which include the following components of genesis indices:

- Potential intensity (PI): the theoretical maximum intensity a TC can reach given the thermodynamic environment (Emanuel 1988; Bister and Emanuel 2002), using reversible ascent from the lowest level, $c_k/c_d = 0.9$, no dissipative heating, and a 0.8 reduction factor to convert the gradient wind to the 10-m wind.
- Column relative humidity: the ratio of the column integrated water vapor path (w) and the column saturated water vapor path (w_s), or w/w_s, following the definition of Bretherton et al. (2004).

- Saturation deficit integrated across the column: the difference between the column integrated water vapor path and the column saturated water vapor path ($w - w_s$).
- Vertical wind shear between 200 and 850 hPa.
- Relative humidity at 600 hPa.
- Absolute vorticity at 850 hPa.
- Vertical motion (omega) at 500 hPa.

We additionally consider outgoing longwave radiation (OLR) since the MSE variance budget [section 2b(2)] considers radiative feedbacks. While genesis indices are designed to predict the number of TCs that form, they are also used as a measure of the general favorability of the environment for TC development (including both formation and intensification), which is the context in which we use them here. We thus relate the genesis indices and large-scale environmental fields to two bulk metrics of TC activity: 1) the climatological mean number of TCs in the peak season (NTC; August–October for the Northern Hemisphere, January–March for the Southern Hemisphere) and 2) accumulated cyclone energy (ACE; Bell et al. 2000) in the peak season. ACE is defined as $\sum v^2$ for all 6-hourly TC snapshots, where v is the maximum sustained 10-m wind speed, and is thus a combination of the number of TCs, their intensity, and their duration.

2) Moist static energy variance budget

The column-integrated MSE $\hat{h}$ is given by

$$\hat{h} = \frac{1}{g}\int_{p_t}^{p_b} (c_p T + gz + L_v q)dp, \tag{1}$$

where p_t is the pressure at the top of the atmospheric column, p_b is the pressure at the bottom of the column, q is the water vapor mixing ratio, and g, c_p, T, z, and L_v have their conventional meanings. The circumflex/hat symbol (ˆ) indicates integration throughout the entire atmospheric column. Moist static energy is approximately conserved under moist adiabatic displacements, and its integral is unchanged by convection if ice processes are ignored. We follow Wing et al. (2019) and integrate over a column with fixed upper and lower bounds of 1 and 925 hPa. This is important so that the same amount of atmospheric mass is integrated over, regardless of the surface pressure of the TC. Wing et al. (2019) noted that an alternate approach is to define a budget for column-integrated MSE per column mass, but we choose the similar and more straightforward option of fixing the pressure bounds. Most TC intensities across each dataset do not fall below 925 hPa (less than 0.5% of all TC snapshots), so we choose this as p_b so that the column integral over the weaker storms is not starting too far above the surface. The very small fraction of snapshots that have surface pressures below 925 hPa would have missing (or interpolated) data in the lowest few levels, which is not expected to impact our results. We note that application of this methodology to other models, which may simulate stronger TCs, may need to consider a different p_b [e.g., Wing et al. (2019) used 920 hPa].

The budget for $\hat{h}$ is given by

$$\frac{\partial \hat{h}}{\partial t} = F_k + N_L + N_S - \widehat{\mathbf{u}\cdot\nabla h}, \tag{2}$$

where F_k is the surface enthalpy flux (the combination of sensible and latent heat fluxes), N_L is the column longwave radiative flux convergence, N_S is the column shortwave radiative flux convergence, and $-\widehat{\mathbf{u}\cdot\nabla h}$ is the column-integrated advection. Each term on the right is a source or sink of $\hat{h}$. The term $\hat{h}$ is calculated using the column integral described above, using instantaneous 3D output at the time of the TC snapshot, and $\partial\hat{h}/\partial t$ is computed using a centered finite difference. The advection term is difficult to calculate accurately from available output, so it is often considered as a residual from the rest of the budget. However, this residual would include more than just the advective term (i.e., the analysis increment and possible imbalances in the budget), so we do not consider it here and instead focus on the diabatic terms. We caution that moist static energy budgets in reanalyses often have imbalances on the order of the terms themselves (e.g., Kiranmayi and Maloney 2011), since reanalyses do not conserve mass or energy (Chiodo and Haimberger 2010; Gelaro et al. 2017). While the individual source terms have been shown to be similar compared to estimates from sounding arrays in some reanalyses (e.g., Sobel et al. 2014), it is important to note the inherent uncertainty in using reanalyses to calculate such budgets, which may imprint on our results. All terms are computed for each grid point (on each reanalysis's own grid) in a 10° × 10° box centered around a TC for each snapshot following its track from TempestExtremes, following Wing et al. (2019). We calculate each of the diabatic flux terms in Eq. (2) as an average centered on the times matching the TempestExtremes tracks for each dataset. To accomplish this, some manipulation of the data was required based on the availability of the radiative and surface fluxes for each reanalysis. For example, the ERA-Interim fluxes were provided as accumulations measured in joules per square meter in 3-h increments from 0000 UTC, so we converted them to averages with units of watts per square meter and centered them around the desired 6-hourly increments at 0000, 0600, 1200, and 1800 UTC to match the TempestExtremes times. We ignore grid points with more than 20% of land as well as grid points poleward of 30° latitude, following prior work (Wing et al. 2019), to focus on the region in which most TCs are found and on the structures of TCs in the tropics. This has the effect of excluding most subtropical storms and storms undergoing extratropical transition, though we do not specifically filter based on the storm classification.

Once the data are extracted in a 10° box centered around each TC snapshot, we calculate the anomalies from the cosine of the latitude-weighted box average of each term. This size box is a compromise between capturing both the TC and its environment and excluding other disturbances and other TCs (Frank 1982; Krouse and Sobel 2010). We test the sensitivity of the results to box size in appendix B. We then compute the budget of the spatial variance of $\hat{h}$, which is given by

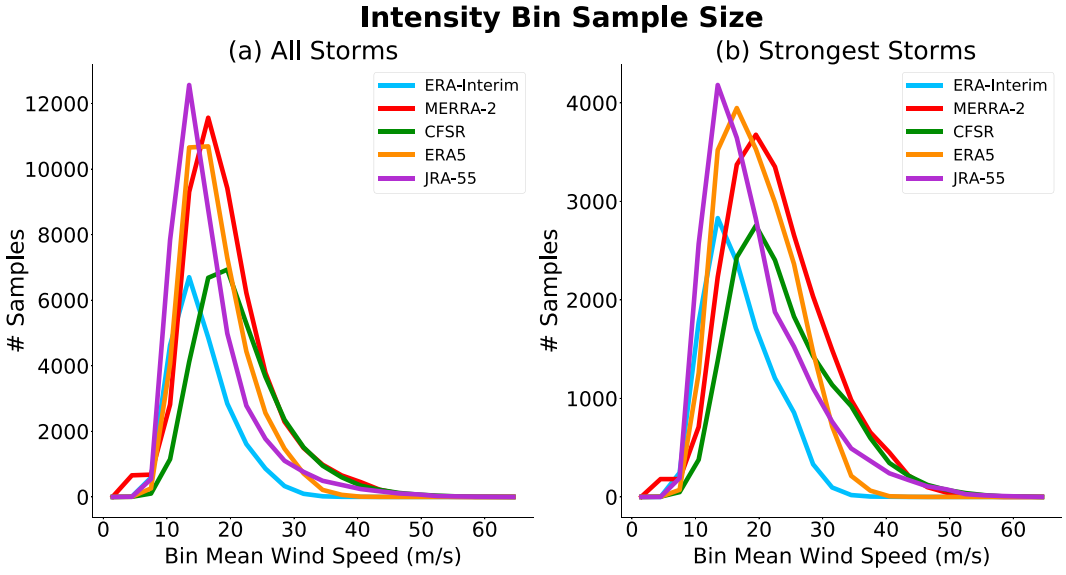

FIG. 3. Comparison of the snapshot sample-size distribution for each intensity bin. Each reanalysis is represented by a different color. (a) Result when all storms from each dataset are included in the composites. (b) The distribution when we restrict the snapshots to only storms whose LMI falls into the 75th percentile of all LMIs across all five reanalyses.

$$\frac{1}{2}\frac{\partial \hat{h}'^2}{\partial t} = \hat{h}'F'_k + \hat{h}'N'_L + \hat{h}'N'_S - \hat{h}'(\widehat{\mathbf{u}\cdot\nabla h})'. \qquad (3)$$

The primes indicate anomalies from the box averages of each individual term from Eq. (2). Equation (3) will be referred to as the MSE variance budget throughout the rest of this paper. Each term in the MSE variance budget contributes to the tendency of MSE variance, where, as the spatial variance of $\hat{h}$ increases, the moist regions of the 10° × 10° box following a TC get moister and the dry regions get drier. $\hat{h}'F'_k$ is the surface flux feedback, $\hat{h}'N'_L$ is the longwave feedback, and $\hat{h}'N'_S$ is the shortwave feedback, also referred to as h'SEF′, h'LW′, and h'SW′, respectively. We focus on these three diabatic feedback terms as sources and sinks of MSE variance and do not consider the advection term of Eq. (3). In this regard, we do not compute a closed budget but rather compare the diabatic feedbacks across reanalyses. The advection term, as well as analysis tendencies from the data assimilation, also likely contribute to the evolution of MSE variance in reanalyses. But the diabatic feedbacks, in the context of the MSE variance budget, are considered to be "process diagnostics," as they represent physical mechanisms relevant to TC development, and thus it is still informative to consider them in isolation as metrics for those processes.

3) INTENSITY-BIN COMPOSITES

To use the diabatic feedbacks in the MSE variance budget described in section 2b(2) as process diagnostics, we construct composites over snapshots of the same TC intensity, following Wing et al. (2019). While Wing et al. (2019) also considered composites relative to the time of lifetime maximum intensity (LMI), we do not show those composites here, as they reveal qualitatively similar results. We bin each snapshot along a TC's track, up through the time of LMI into 3 m s^{-1} increments of maximum near-surface wind speed. After the snapshots are individually binned, we can take the composite average of the terms in Eq. (3) over each bin. This approach allows us to compare storm snapshots of similar intensity. Figure 3a shows the number of individual snapshots across all TCs filed into each 3 m s^{-1} bin. From this plot, we can infer that analyzing the terms in Eq. (3) in the intensity bins between about 10 and 30 m s^{-1} will provide a good sample size (thousands of snapshots) over which to composite. When considering composites of feedbacks averaged over the TC-following box (section 4c), we follow Wing et al. (2019) and take the box average of the feedbacks over the innermost 5° × 5° around the TC center, to focus on the contributions of the different feedbacks near the TC center and of the strongest storms (those whose LMI falls in the 75th percentile of all LMIs within each individual dataset), as those storms experience substantial intensification and are the ones for which the MSE variance budget is thought to be the most relevant. Note that the feedbacks are still calculated from anomalies with respect to the full 10° × 10° box; we are just averaging the feedbacks over a smaller segment of the box. While considering only the strongest storms reduces the sample size in our composites, there are still thousands of samples in the bins between 10 and 30 m s^{-1} (Fig. 3b). Averages over the full 10° × 10° box and over all storms reveal qualitatively similar behavior (see Figs. S8 and S9).

3. Large-scale environment

a. *Climatological mean reanalysis environments*

As shown in section 2, the TC climatology including both the number of TCs and their intensity distribution differs

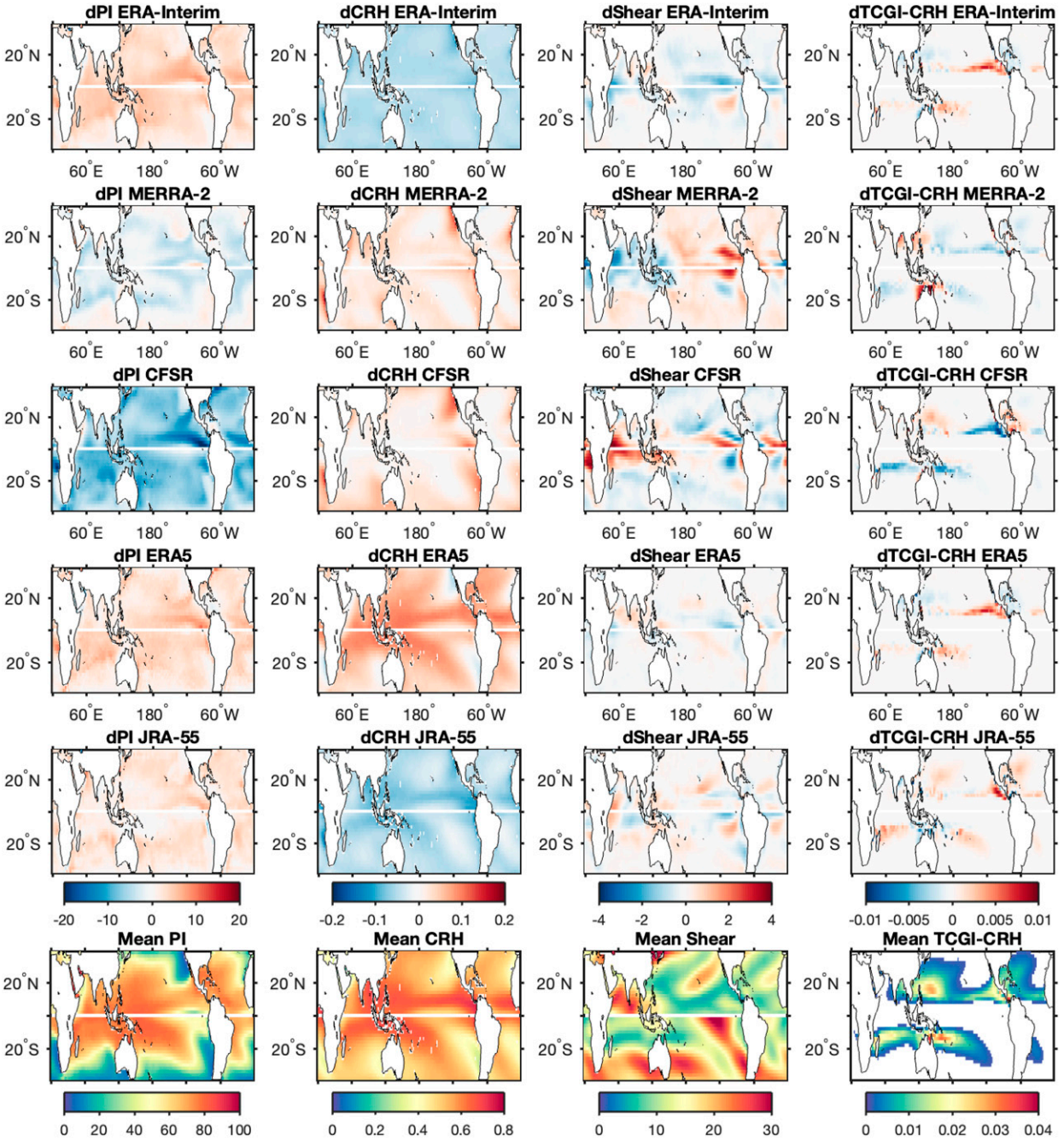

FIG. 4. Climatological mean environmental fields, averaged over 1980–2016 and August–October in the Northern Hemisphere and January–March in the Southern Hemisphere. (bottom) Mean across all reanalyses for (left to right) PI (m s^{-1}), column relative humidity (CRH), 850–200-hPa shear (m s^{-1}), and the TCGI using CRH. (remaining rows) Difference of each reanalysis from the mean across all reanalyses.

across reanalyses. One possible cause for this discrepancy is that large-scale environmental variables that are favorable to TC development are represented differently across reanalyses' mean states. The inter-reanalysis mean TCGI captures the observed spatial pattern of TCs (Fig. 4), but there are differences across reanalyses. For example, MERRA-2 and CFSR depict lower values of TCGI in the eastern Pacific while ERA-Interim, ERA5, and JRA-55 have higher values. This suggests that the environment is more favorable for TC development in the eastern Pacific in the latter three reanalyses. Other versions of genesis indices look similar (Fig. S1). There are also differences in individual components of the

genesis indices; for example, ERA-Interim and JRA-55 have uniformly lower column relative humidity than the inter-reanalysis mean, and CFSR has lower PI (Fig. 4). There is more spatial variability in the differences of vertical wind shear from the inter-reanalysis mean, which reflects differences in atmospheric circulation patterns. The variability across reanalyses in saturation deficit and relative humidity at 600 hPa is similar to that in column relative humidity (Fig. S1). The inter-reanalysis mean OLR depicts regions of deep convection in the intertropical convergence zone (ITCZ) and western Pacific warm pool (low values of OLR) and strong cooling to space in the dry subtropical regions (Fig. S2). Differences in OLR across reanalysis reflect both shifts in the location of the ITCZ and differences in radiative cooling; for example, MERRA-2 generally has much lower values of OLR, whereas JRA-55 generally has larger values. The mean-state 850-hPa vorticity and 500-hPa omega are similar across reanalyses (not shown). Relative to the mean values, hemispheric-averaged saturation deficit has the largest spread across reanalyses, whereas vertical wind shear has the lowest spread. The differences in these TC-relevant, large-scale environmental variables across reanalyses are perhaps not surprising given that some datasets have known biases in certain variables. For example, there are known stratospheric temperature biases in ERA-Interim and ERA5 (Hersbach et al. 2020; Simmons 2020) and a cold bias below 925 hPa in ERA5 that influences the convective instability (Slocum et al. 2022); this could contribute to ERA5's high bias in column relative humidity. However, despite the differences, the climatological large-scale environments represented by reanalyses are much more similar to each other than the environments simulated by GCMs that are unconstrained by observations (Camargo et al. 2020).

b. Relationship with TC climatology

Despite their differences, the climatological large-scale environments do not have a consistent relationship with climatological TC activity across reanalyses. While it is difficult to assess statistical significance with only five datasets, there is a tendency for reanalyses with greater average relative humidity at 600 hPa to simulate more TCs and higher ACE, though this is not universally true and is not found in the other humidity-related variables (Figs. 5a–c and 6a–c). In some cases (i.e., shear, vorticity, and omega), the relationships between the environmental variable and NTC or ACE are different in the Northern versus Southern Hemisphere (Figs. 5e–g and 6e–g). There are also some relationships that are not physically consistent: higher potential intensity and lower shear appear to be associated with lower ACE in the Northern Hemisphere, which is the opposite of expectations (Figs. 5d,e and 6d,e). Reanalyses with more TCs in the Southern Hemisphere tend to have lower average values of OLR, which may reflect, overall, more deep convective activity in those datasets (Fig. 5h). Overall, discrepancies in the climatological large-scale environments do not clearly explain the differences in TC climatologies across reanalyses, nor does model resolution (higher resolution reanalyses tend to have fewer storms, which is opposite of expectations; see Fig. S3). This conclusion is qualitatively unchanged if individual basins, such as the North Atlantic, western North Pacific, or South Pacific, are considered rather than whole hemispheres (not shown).

These results emphasize that simply considering mean-state variables is not sufficient to explain differences in the TC climatology across reanalyses, which motivates considering the physical processes of TC development. This could include how synoptic-scale environments impact storm-level processes; indeed, recent work found that kinematic and thermodynamic variables in the TC environment are biased low in ERA5 relative to dropsonde data (Slocum et al. 2022). In this study though, we utilize diagnostics that focus on storm-scale processes related to how convection, moisture, clouds, and circulation are coupled. The analyses to follow that utilize the MSE variance budget reflect processes occurring at the convective scale within the storm as well as those in the storm's surrounding environment.

4. MSE variance budget

First, we present the spatial structure, magnitude, and intensity dependence of the MSE variance budget across the reanalysis datasets and examine what features the reanalyses agree on and where the reanalyses differ. We then attempt to relate the diabatic feedbacks in the reanalyses to aspects of their TC climatologies (i.e., mean TC intensity, NTC, and ACE).

a. 2D spatial structure

The 2D spatial structure of the MSE and diabatic feedbacks, shown at each grid point in the $10° \times 10°$ box following the TCs, depicts how each reanalysis represents the spatial structure of TCs. As a representative example, we show here the 24–27 m s^{-1} bin (Fig. 7), but similar figures for other bins are provided in the supplemental material (Figs. S4–S7). In subsequent sections, we consider azimuthal and box averages of the feedbacks across all intensity bins. The reanalyses generally depict similar spatial structures, with the exception of ERA-Interim's surface flux feedback term (second column, bottom row of Fig. 7). CFSR and JRA-55 simulate the smallest $\hat{h}$ values in the center of the TC, even at storm snapshots of similar intensity, which could indicate lower humidity or weaker TC warm cores in those datasets. $h'\text{SEF}'$ is negative near the center where surface fluxes are reduced due to weaker winds. The area of negative $h'\text{SEF}'$ in the center is much larger in ERA-Interim than other datasets, spanning hundreds of kilometers. This feature imprints on both the box and azimuthal average analyses (see sections 4b and 4c) to such a degree that the average feedback becomes negative. $h'\text{SEF}'$ is directly correlated to the strength of the wind speed, so we consider a composite of the 10-m wind speed following the same methodology as in section 2b(3). Fig. 8 reveals an incredibly broad 10-m wind field, in which the radius of maximum winds simulated by ERA-Interim is on the order of hundreds of kilometers [also seen in Schenkel et al. (2017)]. The very large, nine-gridpoint area of weaker-than-average wind speeds corresponds to an area of suppressed surface

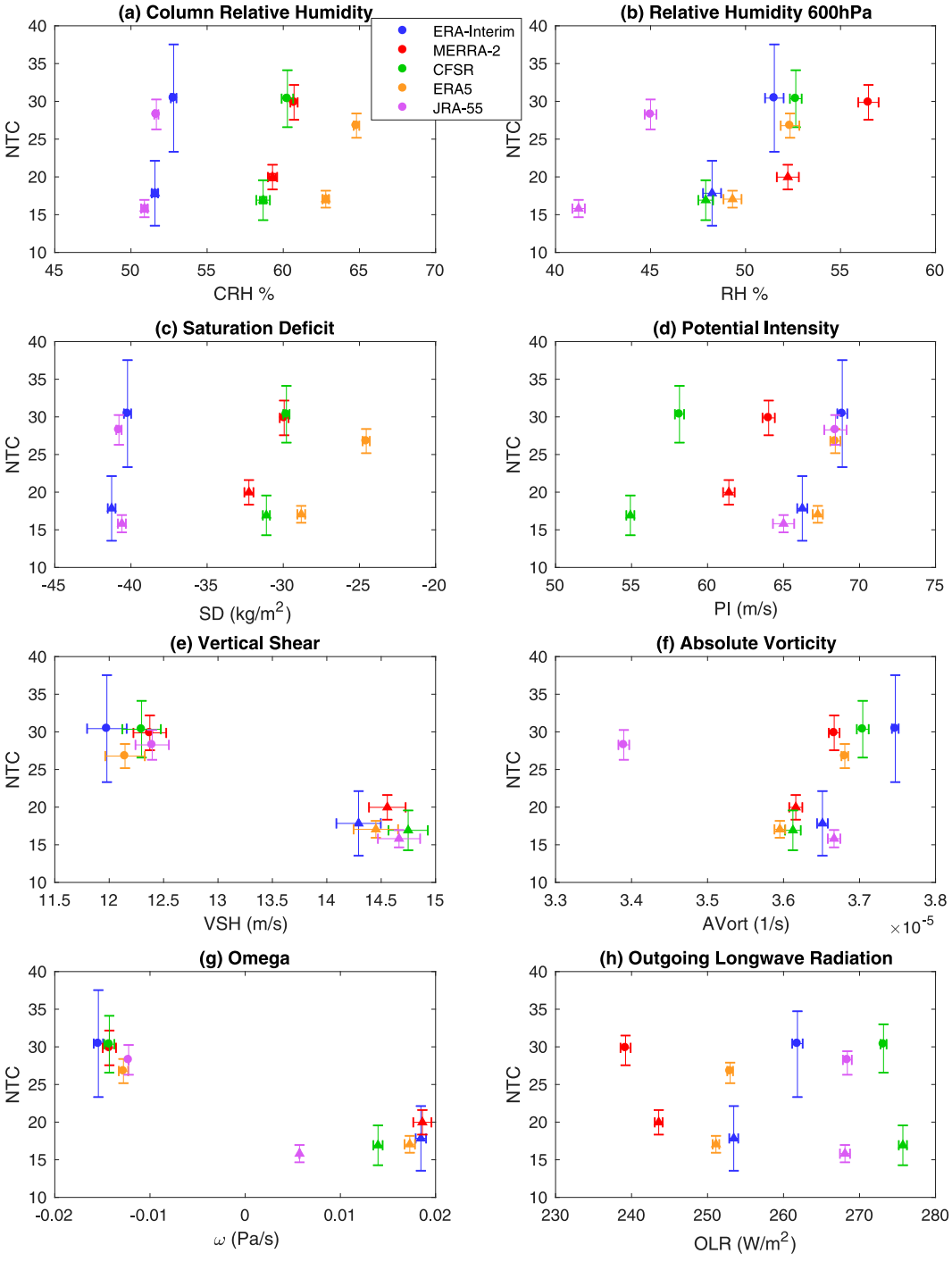

FIG. 5. Scatterplot of the climatological NTCs and mean environmental fields averaged over ocean areas equatorward of 30°. Each is accumulated or averaged, respectively, over August–October in the Northern Hemisphere (circles) and January–March in the Southern Hemisphere (triangles). The error bars indicate the 5%–95% confidence interval for NTCs and the environmental fields, assuming a standard normal distribution.

fluxes (a negative SEF anomaly). Since this is collocated with the largest $\hat{h}$ values ($\hat{h}' > 0$), it results in a negative h'SEF′.

h'LW′ is generally positive, due to reduced longwave cooling in regions of deep convective clouds near the TC center, and is largest in MERRA-2, JRA-55, and ERA5. The large h'LW′ feedback in MERRA-2 (see also Fig. 9b) is consistent with the results of Wing et al. (2019), who examined TCs in the Goddard Earth Observing Systems (GEOS) model

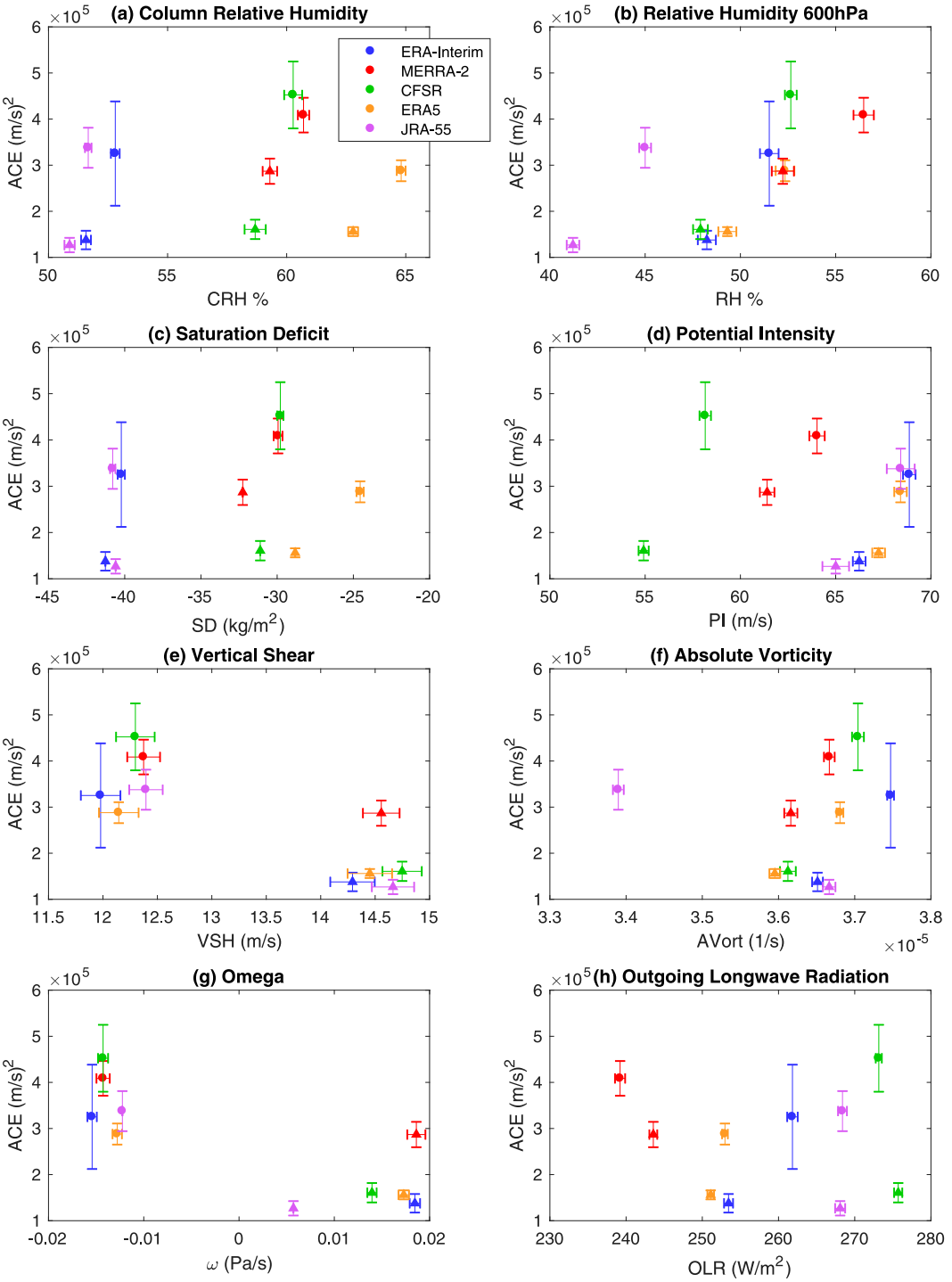

FIG. 6. Scatterplot of the climatological mean ACE and mean environmental fields averaged over ocean areas equatorward of 30°. Each is accumulated or averaged, respectively, over August–October in the Northern Hemisphere (circles) and January–March in the Southern Hemisphere (triangles). The error bars indicate the 5%–95% confidence interval for ACE and the environmental fields, assuming a standard normal distribution.

(Rienecker et al. 2008; Molod et al. 2015), the same model used in MERRA-2. $h'SW'$ is typically weakly positive, due to enhanced shortwave heating in moist areas, except for in ERA5, in which it is weakly negative. We speculate that this could be due to a larger contribution from clouds (which generally reduce the atmospheric shortwave heating by blocking solar radiation) in ERA5. $h'SW'$ generally is the smallest magnitude of the diabatic feedbacks, whereas $h'SEF'$ is

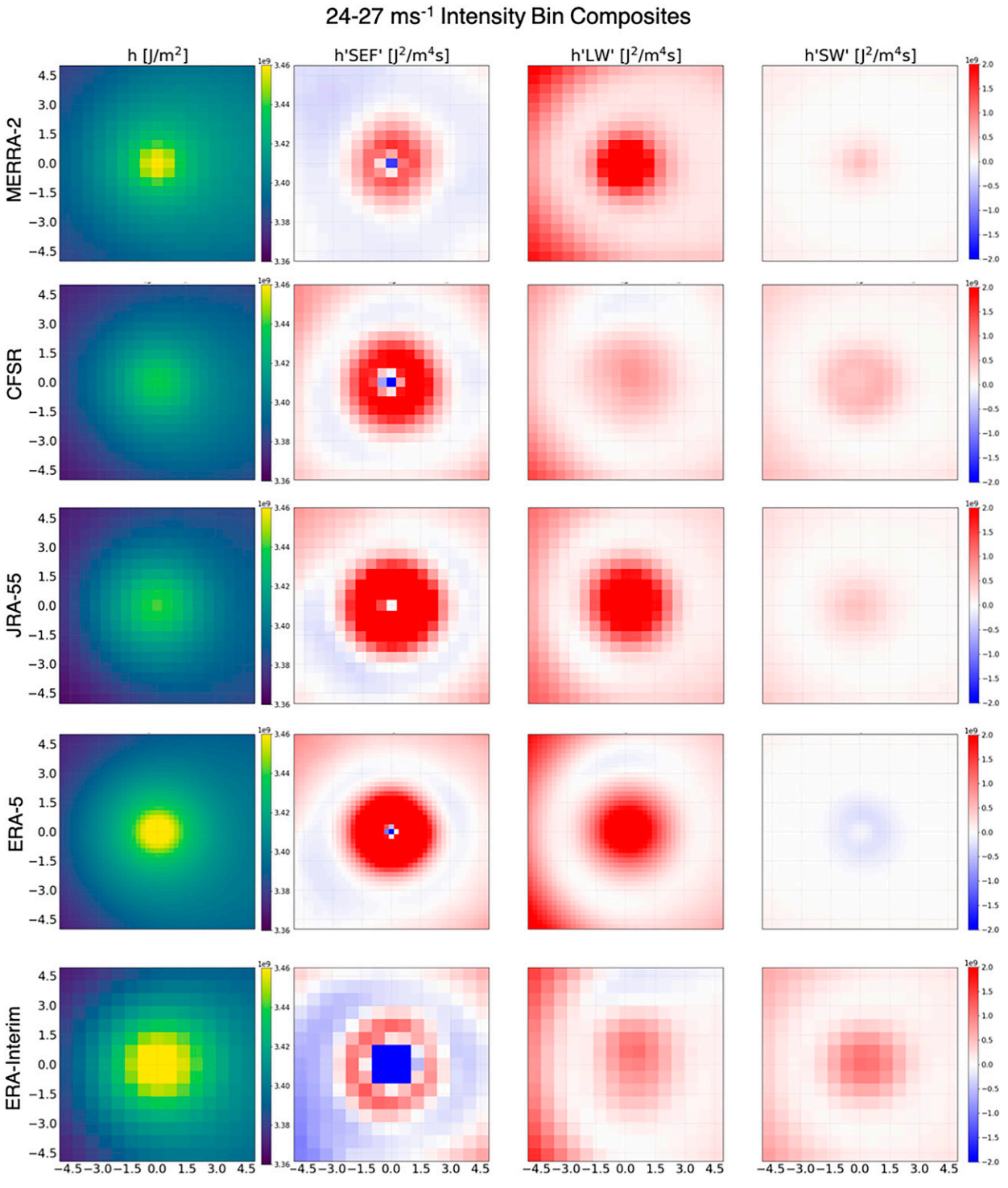

FIG. 7. A 2D spatial view of composites for the 24–27 m s^{-1} bin. Each row represents a different reanalysis. All terms are plotted as a function of degrees from the composited TC center. (left) $\hat{h}$ (J m^{-2}) and (remaining columns) Feedback terms in the MSE variance budget ($\hat{h}'F_k'$, $\hat{h}'N_L'$, $\hat{h}'N_S'$, or h'SEF′, h'LW′, and h'SW′, respectively, with units of J^2 m^{-4} s^{-1}). For the feedback terms, positive values indicate a source of MSE variance, while negative values indicate an MSE variance sink.

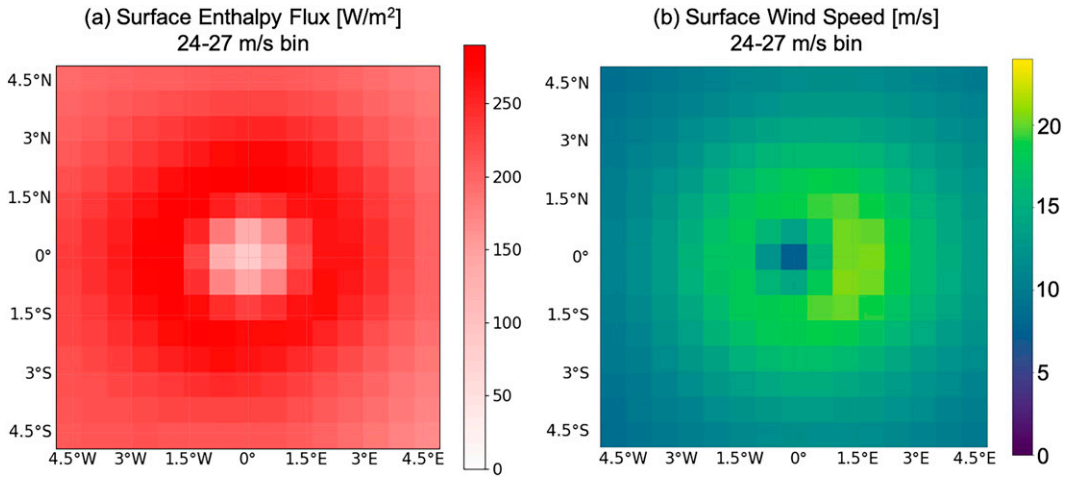

FIG. 8. (a) Surface enthalpy flux and (b) 10-m wind speed composite for ERA-Interim in the 24–27 m s^{-1} bin.

generally the largest and, thus, most important in this intensity bin.

b. *Azimuthal means*

To compute the azimuthal mean from the 2D composites, we interpolate the gridpoint data onto polar coordinates and take the azimuthal average for a given radius out from the center of the TC. Despite some similarities among datasets, there are differences in the value and radial structure of the feedbacks. Figure 9 shows that h'SEF$'$ is smallest, and sometimes negative, closest to the center of the storm and becomes large and positive within 1° of the TC center. Since MSE peaks at the center of the composite TCs (Fig. 7), the radial structure of h'SEF$'$ is driven by the radial structure of the surface enthalpy fluxes, which, in turn, is driven mostly by the radial structure of the surface wind field (Fig. 8). ERA-Interim is an outlier, with a strong negative h'SEF$'$ for most of the radii. We speculate that the negative h'SEF$'$ resulting from the broad wind field (as discussed above) could contribute to weaker storms in ERA-Interim (Fig. 3), since this reanalysis is missing what is typically a strong positive feedback on TC intensification. Some reanalyses simulate a stronger rate of feedback increase with TC intensity than others. For example, the ERA5 h'SEF$'$ magnitude increases with intensity at a much higher rate than MERRA-2's h'SEF$'$, denoted by the vertical spacing between lines of different thickness in Fig. 9a. h'SEF$'$ peaks at different radii (Fig. 9a) for each reanalysis, which likely reflects differences in horizontal resolution and resulting differences in the radius of maximum wind and the structure of the wind field. Especially close to the center and within the radius of maximum winds, h'SEF$'$ has a spread of $O(10^{10})$ J^2 m^{-4} s^{-1} across the datasets in the upper intensity-bin range, while h'LW$'$ and h'SW$'$ have a spread closer to $O(10^9)$ J^2 m^{-4} s^{-1}. For the radiative feedbacks in particular

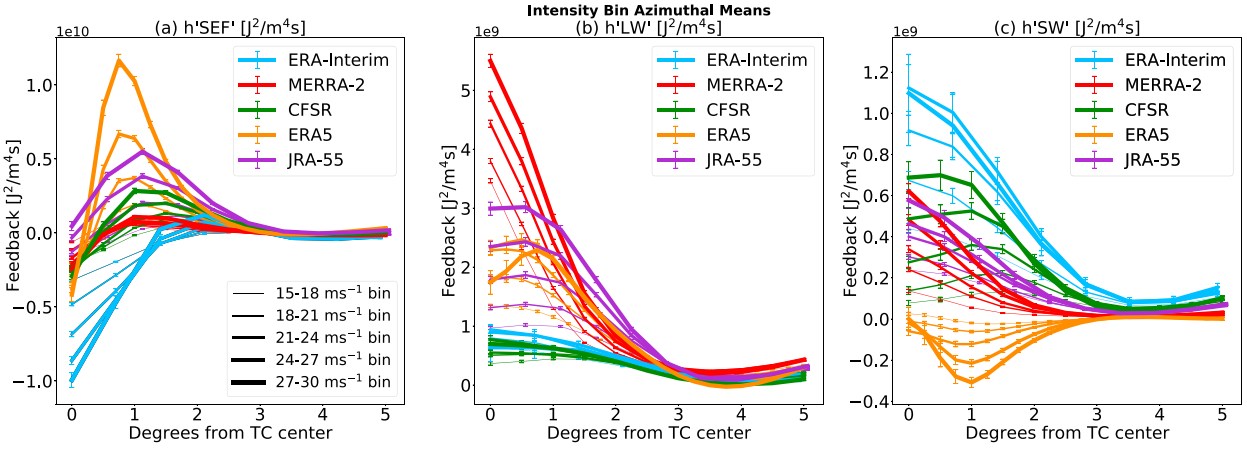

FIG. 9. Azimuthal averages of the intensity-bin composites of the MSE variance budget feedback terms. The different colors represent each reanalysis and each line thickness is a different intensity bin for the (a) surface flux feedback, (b) longwave feedback, and (c) shortwave feedback. Note the different y axes. The error bars indicate the 5%–95% confidence interval, assuming a standard normal distribution.

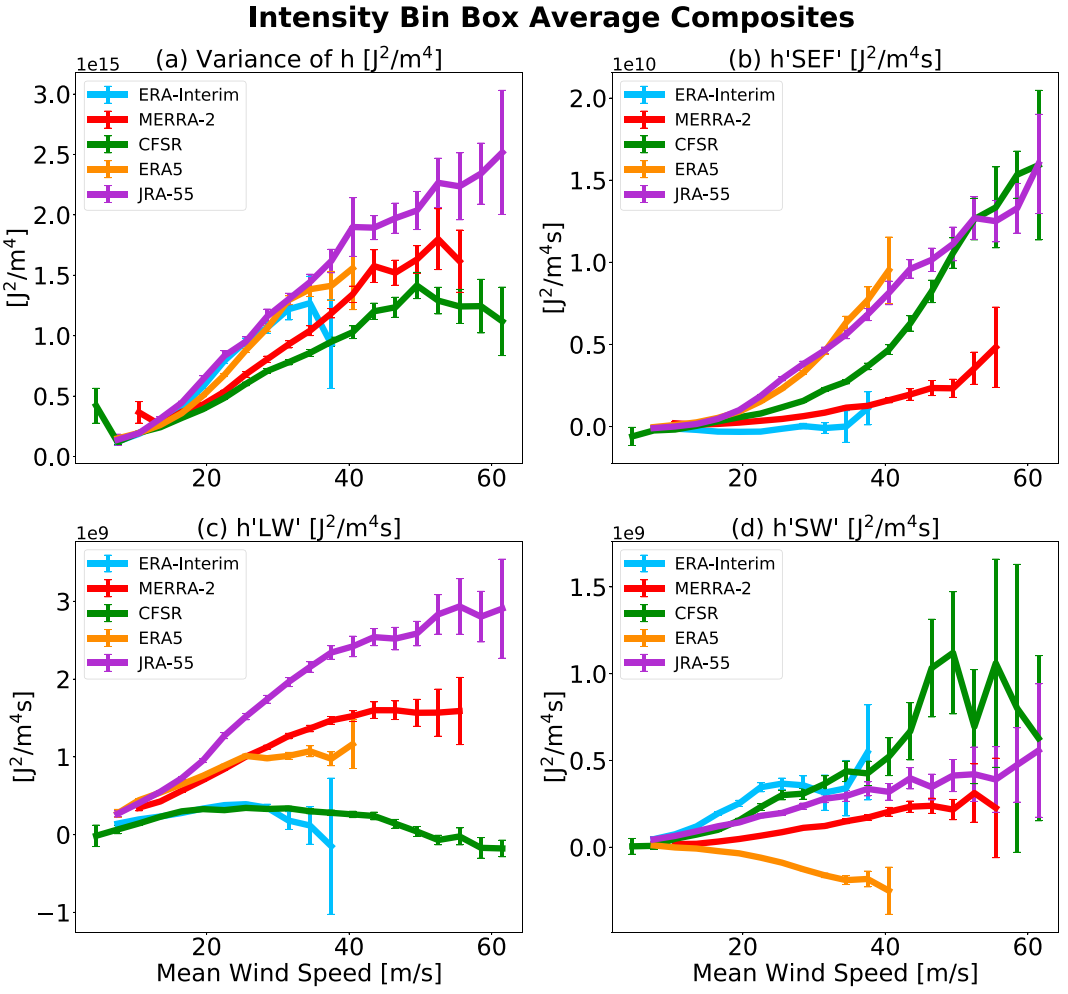

FIG. 10. MSE variance budget feedback terms averaged over the innermost 5° × 5° of the box and composited over the strongest storms (75th percentile of LMIs). Note the different *y* axes. The error bars indicate the 5%–95% confidence interval, assuming a standard normal distribution.

(Figs. 9b,c), the reanalyses exhibit different radial gradients, that is, the rate of change with distance from the TC center. MERRA-2 has the steepest radial gradient for $h'\mathrm{LW}'$, while CFSR has the shallowest. For $h'\mathrm{SW}'$, ERA-Interim has the steepest radial gradient, and CFSR has the shallowest. These differences in the radial gradients of the feedback terms reflect differences in how surface and radiative fluxes vary spatially relative to the variability of the MSE as well as the size of the storms, the latter of which is influenced by horizontal resolution. Consistent with Fig. 7, the azimuthal mean $h'\mathrm{SW}'$ is negative in ERA5, with a peak around 1° from the TC center. Finally, we note that the azimuthal mean of each feedback asymptotically approaches zero near 4° from the TC center, which implies that the relevant processes causing differences in the feedbacks are found mostly inside the TC, not the outer environment.

c. *Box averages*

Prior work has shown that the box-average MSE variance increases with TC intensity in climate models (Wing et al. 2019); here, we show that this relationship is also found in reanalyses (Fig. 10a). As a storm increases in intensity, it moistens compared to its surroundings, thus increasing the MSE variance across the TC. The box-average MSE variance increases roughly linearly with intensity, more than quadrupling in value between the 10 and 40 m s^{-1} intensity-bin composites (Fig. 10a). For a given wind speed, there is a large spread in the values of MSE variance simulated by each reanalysis, which could be due to differences in the strength of the TC warm core, the moisture within the TC, or the environmental moisture. For example, in the 24–27 m s^{-1} bin, the relative spread in the MSE variance, which is defined as the range across reanalyses for a given intensity bin divided by the inter-reanalysis mean for that bin, is 0.43. This indicates that the range across reanalysis is ~40% of the mean value. There are also different rates of increase of MSE variance per increase in wind speed (slopes of lines in Fig. 10a). JRA-55, ERA-Interim, and ERA5 have the largest rate of change in variance for a given change in wind speed, whereas CFSR has the smallest. The differences in these slopes reflect

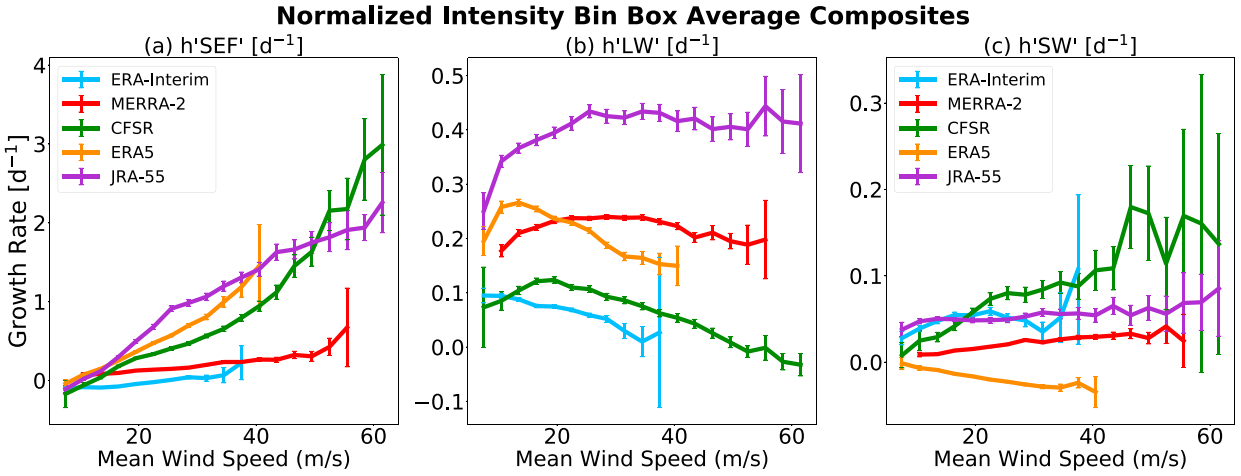

FIG. 11. MSE variance budget feedback terms normalized by the domain mean variance, averaged over the innermost 5° × 5° of the box, and composited over the strongest storms (75th percentile of LMIs). Note the different y axes. The error bars indicate the 5%–95% confidence interval, assuming a standard normal distribution.

different relationships between MSE variance and TC intensity in the different datasets.

As the mean wind speed increases, the values of each feedback also increases. This makes sense particularly for h'SEF′, because the surface enthalpy flux is directly dependent on the wind speed. As deep convection intensifies and the clouds become more concentrated with increasing TC intensity, the radiative feedbacks generally increase in magnitude as well. The box average of h'SEF′ in ERA-Interim (Fig. 10b) is negative at all intensities and does not vary with intensity. JRA-55, MERRA-2, and ERA5 have the largest box-average h'LW′ across most intensities, and the strongest increase with intensity. ERA-Interim and CFSR exhibit a much weaker h'LW′ that does not change much with intensity. Consistent with Figs. 7 and 9, ERA5 has a negative box-average h'SW′ at all intensities that increases in magnitude as intensity increases. Despite the agreement in their qualitative trends with intensity, the actual value of each box-averaged feedback varies statistically significantly across the reanalyses for a given wind speed. Note that the y axes for both h'LW′ and h'SW′ are not the same as for h'SEF′ in Fig. 10, but the spread is still on the order of 10^9 J^2 m^{-4} s^{-1} for the midrange-intensity radiative feedbacks. The relative spread is largest for h'SW′, with an average of relative spread of 2.00 (indicating a range that is twice as large as the inter-reanalysis mean) compared to 1.48 for h'LW′ and 1.36 for h'SEF′.

d. *Normalized composites*

As discussed above, the strength of the diabatic feedbacks generally increases as TC intensity increases. This may be partially driven by the increase in $\hat{h}$ anomalies themselves with intensity. Furthermore, as seen in Fig. 10a, at a given intensity there is a fairly large spread in the value of the $\hat{h}$ variance and thus $\hat{h}'$ anomalies across the reanalyses. This implies that even though we are comparing storm snapshots of similar strength, if a particular reanalysis happens to have greater $\hat{h}$ anomalies associated with a particular intensity, that may contribute to larger feedbacks in that reanalysis for the same intensity, as the feedbacks depend explicitly on $\hat{h}'$. To determine whether the different feedback magnitudes are a result of differences in the physical processes or just different baseline $\hat{h}$ anomalies, and to better compare the feedback magnitudes across different intensities, we normalize the feedbacks by dividing them by the domain mean $\hat{h}$ variance of each snapshot before taking the composites [following Wing and Cronin (2016) and Wing et al. (2019)]. This approach scales the feedbacks and allows a more direct comparison across different values of $\hat{h}'$.

The normalized feedbacks (Fig. 11) have units of days^{-1}, representing h'SEF′, h'LW′, and h'SW′ as growth rates of $\hat{h}$ variance per day. h'SEF′ still increases with intensity even after being normalized by the variance, though not quite as prominently as in Fig. 10b. On the other hand, the normalized h'LW′ (Fig. 11b) either decreases or remains constant with TC intensity. A decrease with intensity is most prominent in ERA5, ERA-Interim, and CFSR, and is found in all reanalyses when considering the average over the full 10° × 10° box over all storms (Fig. S9). This implies that h'LW′ is most important for weaker storms and, in fact, for intensities below ~20 m s^{-1}, the total radiative feedback (sum of h'LW′ and h'SW) is larger than h'SEF′. The normalized h'SW′ has the largest relative spread in the 24–27 m s^{-1} bin (2.87) compared to normalized h'SEF′ and h'LW′ (2.12 and 1.79, respectively). The normalized h'SEF′ feedback seems to cluster into two groups of reanalyses: ERA-Interim and MERRA-2 have values that are notably lower at a given intensity and increase with intensity at a slower rate than those in CFSR, ERA5, and JRA-55. Overall, all three feedbacks are represented differently in each reanalysis even when normalized by box-average $\hat{h}$ variance, indicating that these physical processes, which are important to TC development, are not simulated equivalently across reanalysis. This may contribute to the differences in the distribution of TCs simulated across reanalyses, which is explored next.

e. Relationship with TC climatology

Here, we explore the relationship between the magnitudes of the diabatic feedbacks across reanalyses and aspects of their respective TC climatologies in an attempt to attribute the differences in TC climatologies across reanalyses to their representation of physical processes. This is the intended use of a process-oriented diagnostic, though such diagnostics are typically applied to compare a model to observations, not to compare different observationally constrained datasets (i.e., reanalyses) to each other. Given that section 3 demonstrated that differences in the representation of TCs in reanalyses are not readily explained by their large-scale environments, and that section 4 demonstrated that there are differences in the radiative and surface flux feedbacks across reanalyses, it seems plausible to try to apply the diagnostic to explain the spread across reanalyses. To do so, we relate the average radiative and surface fluxes within a given intensity bin (between v_{lower} and v_{upper}) to four metrics of the TC climatology: the average LMI across all storms, the climatological average NTC, the climatological average ACE, and the percentage of storms that reach intensities beyond the intensity bin under consideration. This last quantity is computed as the ratio of the number of storms with LMI greater than v_{upper} to the number of storms with LMI greater than v_{lower}, since the former at some point must have passed through the bin defined by v_{lower} and v_{upper} and then intensified beyond it.

Figure 12 shows the relationships for the nonnormalized and normalized feedbacks averaged over the innermost 5° × 5° of the box and composited over the strongest storms (75th percentile of LMIs), for the 24–27 m s^{-1} bin; other bins are shown in Figs. S10–S14. There are not many clear relationships between metrics of the TC climatology and the average radiative and surface flux feedbacks, and it is difficult to assess statistical significance with only five datasets to consider. The clearest relationship is that there tends to be a stronger longwave radiative feedback in the 24–27 m s^{-1} bin in reanalyses in which there is a greater percentage of storms that intensify from 24 m s^{-1} to greater than 27 m s^{-1}. This indicates that if a particular reanalysis has a stronger longwave feedback at a given intensity, it may be more likely for the storms to intensify further. This relationship is seen in Fig. 12 (see open circles in Fig. 12c,d, as well as in Figs. S13 and S14), though CFSR is an exception to the relationship. Other possible relationships include a tendency for there to be higher surface flux feedbacks in reanalyses where more storms intensify further in some bins (Fig. 12 and Fig. S13). There do not appear to be consistent relationships between the feedbacks in any bin and LMI, NTC, or ACE, though the lack of relationship with NTC and ACE (which depends partially on NTC) might be because here we are considering feedbacks after the TC has already formed. Overall though, there is not a particularly strong indication that the differences in the representation of the TC climatology can be explained by differences in the representation of radiative and surface flux feedbacks across reanalyses. This suggests that the development and intensification of TCs in reanalyses might be more related to their analysis tendencies from the data assimilation, rather than physical process representation, though it could also mean that other physical processes beyond those examined here need to be considered.

5. Discussion and conclusions

We apply the column-integrated MSE variance budget, first introduced by Wing and Emanuel (2014), and as first applied to TCs in global models in Wing et al. (2019), to five global reanalysis datasets to investigate their representation of physical processes involved in TC development and to consider whether reanalyses can serve as an observation-based reference for these diagnostics. This effort is challenged by known deficiencies in reanalysis representation of TCs, but no other dataset provides all the necessary variables.

We first compare climatological large-scale environmental variables associated with TCs as a possible source of spread across reanalyses. The largest relative differences across all reanalysis products occur in variables involving humidity, while the spread is relatively smaller for dynamical variables (e.g., vertical wind shear, omega). However, differences in the climatological large-scale environments do not readily explain the inter-reanalysis spread in the TC climatology; physical process representation at smaller scales must also be considered.

We use the MSE variance budget to diagnose surface flux and longwave and shortwave radiative feedbacks in TCs, which provides a measure of diabatic processes known to be important to TC development. Our results reveal that there is a large spread in the magnitude of these feedbacks, as well as in the MSE variance itself, for a given TC intensity across the reanalyses. All terms generally increase in magnitude as the TC intensifies, but the spread in MSE variance at a given intensity is ~40% of the value of the inter-reanalysis mean value, and the MSE variance increases with intensity at different rates. Even after normalizing by the box-average MSE variance before compositing $h'\text{SEF}'$, $h'\text{LW}'$, and $h'\text{SW}'$ across the TC snapshots, the reanalyses still exhibit a large spread in the values of each feedback, with average relative spreads of 0.85, 1.72, and 2.15, respectively. The normalized surface flux feedback still increases with TC intensity, but the contribution from the normalized longwave and shortwave feedbacks depend less on intensity. Reanalyses agree that the radiative feedbacks have a greater relative contribution to TC development in weak storms, and their sum is, in fact, greater than the surface flux feedback at low intensities (roughly below 20 m s^{-1}).

We conclude that reanalyses have substantial differences ingrained in their representation of physical processes associated with TC development, as captured by the MSE variance budget. It is possible that this could contribute to the variability across reanalyses in their representation of the TC climatology. We found some suggestion that a stronger longwave radiative feedback in a given intensity bin is associated with a higher percentage of storms that intensify further, consistent with the growing body of literature on the importance of radiative feedbacks for TC development (e.g., Wing 2022; Ruppert et al. 2020; B. Zhang et al. 2021; Wu et al. 2021, 2023). However, the relationship between the diabatic feedbacks and aspects of the TC climatology across reanalyses was not particularly clear,

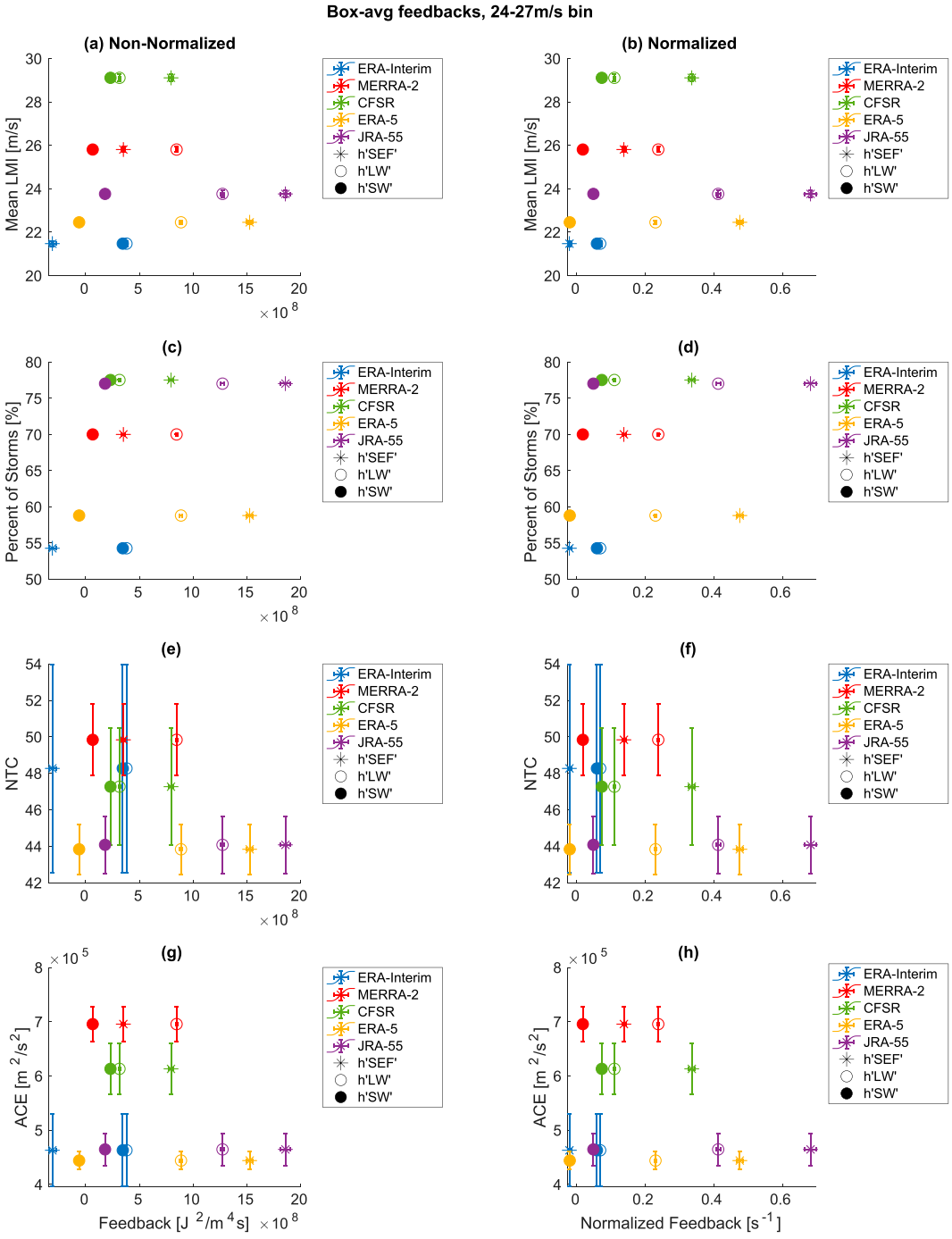

FIG. 12. Scatterplot of (a),(b) the mean LMI across all storms; (c),(d) the percentage of storms intensifying from 24 m s^{-1} to greater than 27 m s^{-1}; (e),(f) the NTCs in the peak season; and (g),(h) the ACE in the peak season against (left) nonnormalized and (right) normalized surface flux (stars), longwave (open circles), and shortwave (filled circles) feedbacks, averaged over the innermost 5° × 5° of the box, and composited in the 24–27 m s^{-1} bin across the strongest storms. Each reanalysis is a different color. The two-way error bars indicate the 5%–95% confidence interval, assuming a standard normal distribution, for the average feedbacks and TC metrics (there are no error bars on the percentage of storms metric; otherwise, if they are not visible, that means that they are very small).

indicating that other factors also contribute to the inter-reanalysis spread in the TC climatology, including other effects on MSE variance not analyzed here, such as the advective term and data assimilation tendencies, as well as other physical processes not captured by the MSE variance budget. Despite this, our results are useful to reanalysis developers since they quantify how each reanalysis depicts surface flux and radiative feedbacks on TCs. The large spread is also a cautionary note to reanalysis users that reanalyses may not be trustworthy in their representation of this budget.

As for what drives the differences in the surface flux and radiative feedbacks, and, more generally, the TC representation, across the reanalyses, there are many potential sources, including different native model resolutions, data assimilation schemes, TC preprocessing techniques, choice of prognostic variables and model equations, and physical parameterizations. All of these aforementioned differences in reanalyses could imprint on the results we see in this study, though horizontal resolution does not seem to explain the differences in the TC climatology across reanalyses. Higher resolution reanalyses tend to simulate fewer TCs and lower ACE (Fig. S3), which is opposite of expectations, and there are also reanalyses at similar resolutions with very different TC climatologies. There also does not appear to be a relationship between the radiative feedbacks and resolution, though $h'\text{SEF}'$ is typically larger in models of higher resolution, though this is not universally true (Fig. S15). Given the myriad differences, it is extremely difficult to attribute to any one factor why, for example, JRA-55 has a stronger surface flux feedback than MERRA-2. Since $h'\text{SEF}'$ is closely tied to the covariability of the wind and moisture fields, if these reanalyses have different boundary layer structures of either of these fields, due to different boundary layer schemes or dynamical cores, or different parameterizations of the wind-speed dependence of surface fluxes, that would imprint on $h'\text{SEF}'$. Boundary layer temperature and moisture biases, such as the known cold bias below 925 hPa in TC environments in ERA5 (Slocum et al. 2022), could imprint on the MSE variance budget, though it is difficult to speculate how since the feedbacks in the MSE variance budget depend not on absolute values of MSE or surface or radiative fluxes but on their spatial variabilities. The different cloud, microphysics, and radiation schemes in each model used to create the reanalyses could influence both the longwave and shortwave feedbacks. But beyond such speculation, attribution would likely require parameter perturbation experiments within the context of an individual model. We must also consider whether the evolution of a TC in a reanalysis product is really driven by physical processes internal to the model or by analysis tendencies from the data assimilation.

Future work will apply these diagnostics to GCMs from phase 6 of the Coupled Model Intercomparison Project (CMIP6) and high-resolution GCMs from HighResMIP (Haarsma et al. 2016). However, the large spread in the MSE variance budget across reanalyses, and their known biases in representing TCs (Hodges et al. 2003; Schenkel and Hart 2012; Murakami 2014; Hodges et al. 2017; Schenkel et al. 2017; Kim et al. 2021; Bian et al. 2021; Zarzycki et al. 2021; Jones et al. 2021) and the TC environment (Slocum et al. 2022), means that we may not be able to consider reanalyses as the "truth" that the GCMs should be evaluated against. The MSE variance budget seems to be relatively unconstrained by the assimilation of observational data (if it was, the reanalyses should agree better), and instead reflects differences in the methods of data assimilation or numerics and physics of the underlying models used to generate reanalyses. A preliminary comparison of the box-average MSE variance and the surface flux and radiative feedbacks in the reanalyses to the six GCMs considered in Wing et al. (2019) indicates that the GCMs fall within the range of the reanalyses, but that the spread across the reanalyses is similar to, if not larger than, the spread across these particular GCMs (not shown). However, the climatological characteristics of TCs in MERRA-2 differ from atmosphere-only runs of the same model without data assimilation (Aarons et al. 2021) and there are also clear differences in the MSE variance budget terms (not shown). This indicates that at least within a given model environment, assimilation of observations does make a difference, though it does not necessarily improve the representation of TCs (Aarons et al. 2021). Future work that evaluates the MSE variance budget for TCs in GCMs against reanalyses must thus be interpreted extremely carefully and should, at minimum, compare to multiple reanalyses and consider the GCMs in the context of the large reanalysis spread. Ongoing efforts to compute the MSE variance budget in observed TCs from a combination of satellite retrievals and dropsonde observations from aircraft reconnaissance are a step toward providing a much-needed true observational reference for this process-oriented diagnostic.

Acknowledgments. This project is supported by NOAA MAPP Grants NA18OAR4310270, NA18OAR4310276, and NA18OAR4310277 and contributes to the NOAA Model Diagnostic Task Force. Daehyun Kim was also supported by KMA R and D program under the grant number KMI2021-01210. MERRA-2 radiative flux, ERA5 radiative flux and surface flux, and JRA-55 radiative flux and surface flux data were downloaded by Evan Jones and Fred Soster. CFSR, CFSv2, ERA-Interim, MERRA-2, and JRA-55 3D data were downloaded by Robert Hart. We thank them, as well as Rhys Parfitt, for providing access to these downloaded data. We also thank Colin Zarzycki for providing the TempestExtremes tracks utilized in this study, Evan Jones for generating an early version of Fig. 1, and Jarrett Starr for identifying a few code errors during the revision process. Caitlin Dirkes also thanks Robert Hart, Rhys Parfitt, and Jeff Chagnon for helpful comments as members of her master's thesis committee. We thank three anonymous reviewers for detailed reviews that significantly improved the manuscript.

Data availability statement. The IBTrACS data were obtained from NOAA's National Climatic Data Center (https://doi.org/10.25921/82ty-9e16; Knapp et al. 2018). Reanalysis

data from CFSR (https://doi.org/10.5065/D6513W89, https://doi.org/10.5065/D69K487J; Saha et al. 2010a,b), CFSv2 (https://doi.org/10.5065/D61C1TXF; Saha et al. 2011), JRA-55 (https://doi.org/10.5065/D6HH6H41; JMA 2013), ERA-Interim (https://doi.org/10.5065/D6CR5RD9, https://doi.org/10.5065/D64747WN; ECMWF 2009, 2012), and ERA5 (https://doi.org/10.5065/BH6N-5N20; ECMWF 2019) were provided by the Research Data Archive (RDA) of the Computational and Information Systems Laboratory at the National Center for Atmospheric Research (NCAR) and are available at https://rda.ucar.edu. NCAR is supported by grants from the National Science Foundation. Reanalysis data from MERRA-2 (https://doi.org/10.5067/7MCPBJ41Y0K6, https://doi.org/10.5067/A7S6XP56VZWS; GMAO 2015a,b), provided by the Goddard Earth Sciences Data and Information Services Center (GES DISC), are available at https://disc.gsfc.nasa.gov. Reanalysis data from ERA5 (Copernicus Climate Change Service 2023) were provided by the Copernicus Climate Change Service Climate Data Store (CDS) and are available at https://cds.climate.copernicus.eu. Calculations with ERA5 and JRA-55 3D data were made with high-performance computing support from Cheyenne (https://doi.org/10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. Reanalysis tropical cyclone tracks are from Zarzycki et al. (2021), derived from the Cyclone Metrics Package (CyMeP; https://github.com/zarzycki/cymep; https://doi.org/10.5281/zenodo.4546658).

APPENDIX A

Sensitivity to Output Resolution

A complicating factor in our analysis is that the different reanalyses have different resolutions, and their output resolution may differ from the native model resolution. While the native model resolution is known to influence TC simulation by influencing which dynamics are resolved (Roberts et al. 2020; Moon et al. 2020b), different output resolutions may also impact the MSE variance budget diagnostics. We investigate this by comparing the JRA-55 intensity-bin composites computed from the native model resolution data (~0.5°) to that computed from data on a 1.25° × 1.25° grid and pressure levels, available for download from NCAR's Research Data Archive.

The spatial structure and magnitudes of the feedbacks are generally similar for the 1.25° and model resolution data, though the large $\hat{h}$ values are more spatially confined in the model resolution data (see the 2D spatial view in Fig. S16). The coarser 1.25° output resolution data cannot capture the inner-core structure and generally underestimates the radial gradient of the feedbacks, with a weaker feedback strength in a given bin at a given radius within ~2.5° of the TC center, but a stronger feedback strength in a given bin at a given radius further outward (Figs. A1a–c). It fails to capture most of the reduction in the $h'\text{SEF}'$ term near the TC center (Fig. A1a). The box-average MSE variance and feedbacks increase similarly with intensity for both the 1.25° output and native model resolution data, but the variance and feedbacks are systematically slightly smaller with the coarser 1.25° data (Figs. A1d–f). However, the difference in magnitude is small compared to the range across reanalyses (Fig. 10). The different output resolutions across reanalyses are thus not expected to be a primary factor in the differences in the MSE variance budget diagnostics. The different model resolutions may, of course, still contribute to these differences.

When comparing the MSE variance budget across models of different resolutions, one might consider first interpolating to a common grid. While this would eliminate output resolution as a source of difference, there is a trade-off. If the diagnostics are computed from coarsened data, they may underestimate the feedbacks that are actually felt by the TCs as they are simulated at native model resolution. We therefore recommend that quantitative comparison of terms in the MSE variance budget is best suited to comparing models that have the same resolution. Another factor that could be explored in the future is the sensitivity of the MSE variance budget to the reanalyses' effective resolutions (resolution that resolves the kinetic energy spectrum), which varies substantially across low-resolution models and is more consistent as the resolution increases (Klaver et al. 2020).

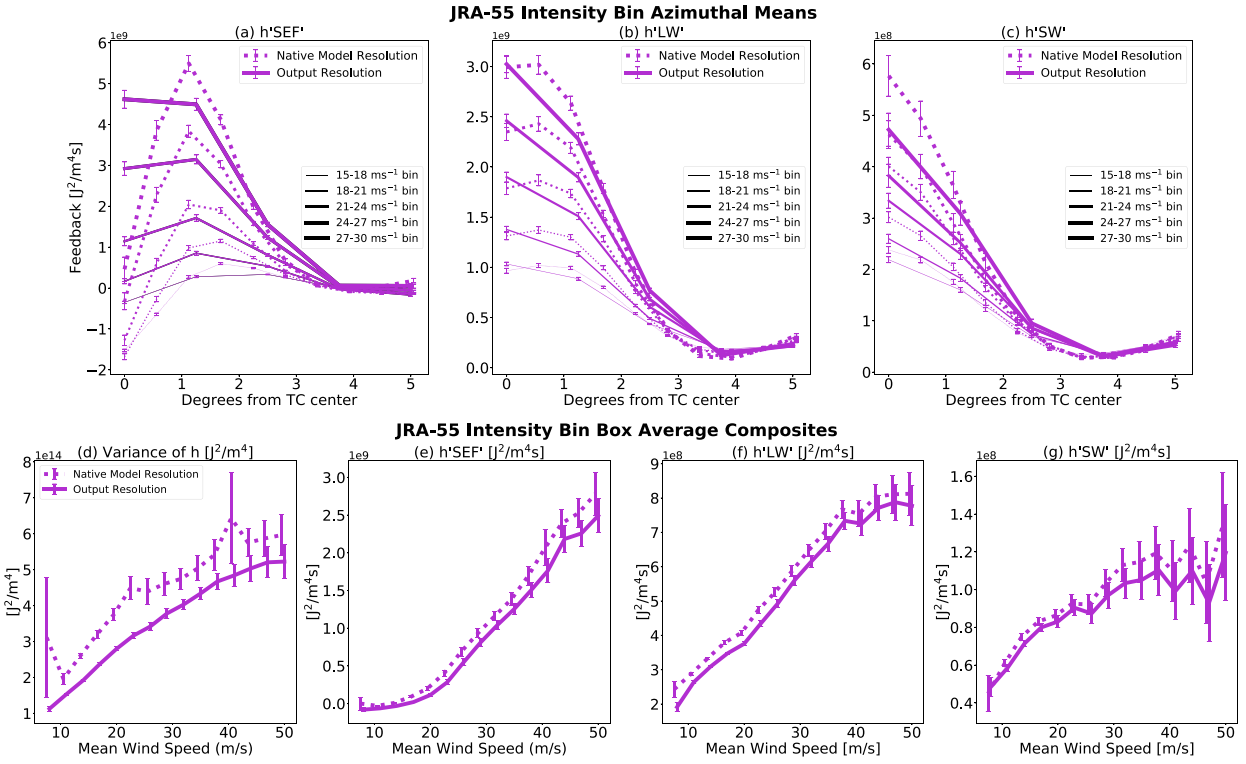

FIG. A1. Sensitivity of MSE variance and feedback calculation to output resolution. (top) Azimuthal averages of the JRA-55 intensity-bin composites of the (a) surface flux feedback, (b) longwave feedback, and (c) shortwave feedback in the MSE variance budget. Each line thickness is a different intensity bin. (bottom) Box averages (over all storms and the full box) of the (d) MSE variance, (e) surface flux feedback), (f) longwave feedback, and (g) shortwave feedback in the MSE variance budget. Note that all feedbacks have different y axes so that the difference between resolutions are visible. The solid lines are the 1.25° output resolution data and dotted lines are the ~0.5° native model resolution data. The error bars indicate the 5%–95% confidence interval, assuming a standard normal distribution.

APPENDIX B

Sensitivity to Box Size

A potential limitation of our results is our choice to calculate the MSE variance budget terms using anomalies from a fixed 10° × 10° box around each TC center. We test the sensitivity of the results to the choice of box size in CFSR and find that whether a 5° × 5°, 10° × 10°, or 15° × 15° box is used to compute the MSE variance budget, the structure of the feedbacks and their dependence on intensity is generally qualitatively similar, but there are substantial differences in their magnitudes. Each feedback term gets larger in magnitude as the box size increases (Fig. B1). For all box sizes, there is a negative h'SEF′ near the TC center, but the area of negative feedback decreases in size with increasing box size (Fig. B1a, also Fig. S17). The azimuthal mean h'SEF′ has a similar radial structure for each box size, with higher peak values in the larger boxes, but the radial gradients of h'LW′ are quite different (Figs. B1a,b). For a given intensity bin, h'LW′ decreases much more quickly with radius for the 15° × 15° box than the 5° × 5° box (Fig. B1b). This is also true to a lesser extent for h'SW′ (Fig. B1c).

The box-average MSE variance still increases with intensity for each box size but at slightly different rates and, at a given intensity, it is larger for larger box sizes (Fig. B1d). Of the box-average feedbacks, h'LW′ exhibits the greatest sensitivity to box size; its magnitude at a given intensity increases with box size and its dependence on intensity differs depending on the box size (Fig. B1e). The ratio of box average h'SEF′ to box average total radiative feedback in the 24–27 m s^{-1} intensity bin for the 5° × 5°, 10° × 10°, and 15° × 15° boxes are 0.70, 1.05, and 0.62, respectively. We expected this ratio to monotonically increase with box size, but it is possible that making the box size either too small or too large causes some cancellation effects between the feedback magnitudes in the TC inner core or outer environment. Thus, the box size must be carefully considered when computing the MSE variance budget. Future work might test the use of a dynamic box size based on resolution or some metric of TC size. However, since most of the reanalyses we use have similar output resolutions, using a fixed box size across all datasets is not expected to significantly impact our results.

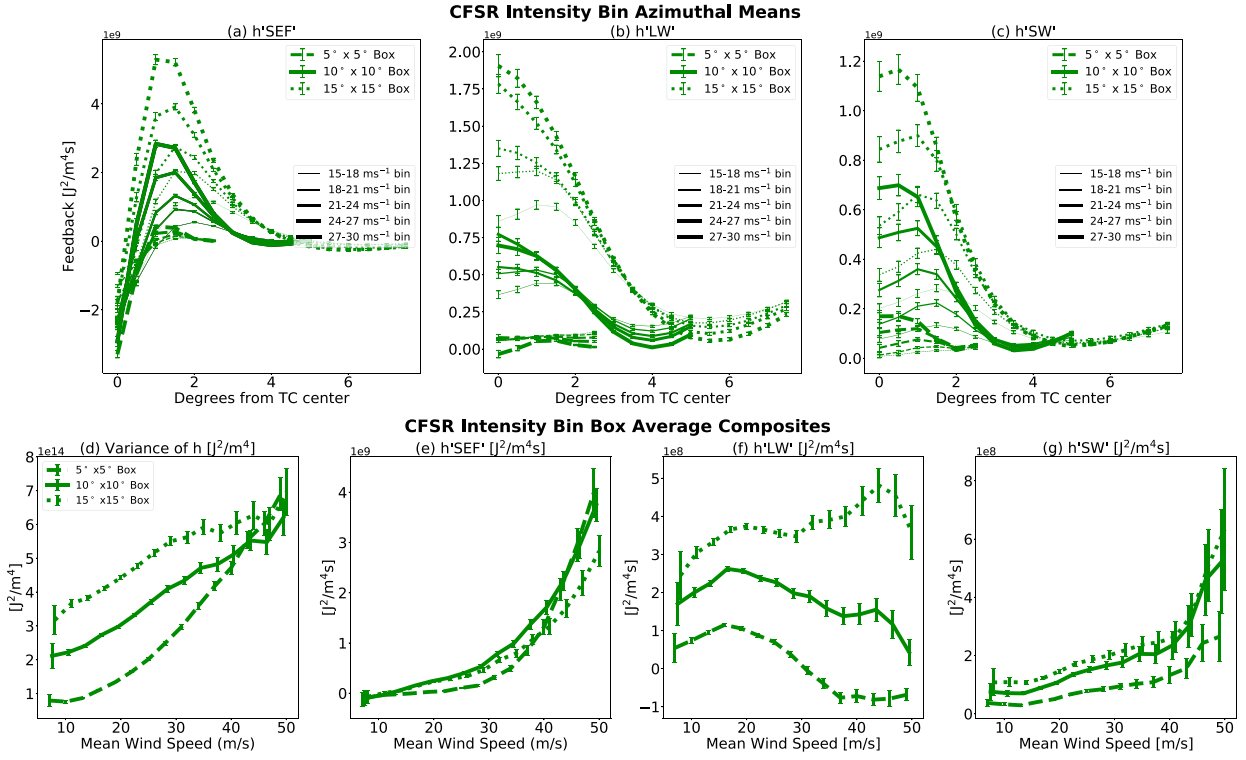

FIG. B1. Sensitivity of MSE variance and feedback calculations to box size. (top) Azimuthal averages of the CFSR intensity-bin composites of the (a) surface flux feedback, (b) longwave feedback, and (c) shortwave feedback in the MSE variance budget. Each line thickness is a different intensity bin. (bottom) Box averages (over all storms and the full box) of the (d) MSE variance, (e) surface flux feedback), (f) longwave feedback, and (g) shortwave feedback in the MSE variance budget. Note that all feedbacks have different y axes so that the difference between box sizes (different line styles) are visible. The error bars indicate the 5%–95% confidence interval, assuming a standard normal distribution.

REFERENCES

Aarons, Z. S., S. J. Camargo, J. D. O. Strong, and H. Murakami, 2021: Tropical cyclone characteristics in the MERRA-2 Reanalysis and AMIP simulations. *Earth Space Sci.*, **8**, e2020EA001415, https://doi.org/10.1029/2020EA001415.

Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment, part I. *J. Atmos. Sci.*, **31**, 674–701, https://doi.org/10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2.

Bacmeister, J. T., M. J. Suarez, and F. R. Robertson, 2006: Rain reevaporation, boundary layer-convection interactions, and Pacific rainfall patterns in an AGCM. *J. Atmos. Sci.*, **63**, 3383–3403, https://doi.org/10.1175/JAS3791.1.

Bell, G. D., and Coauthors, 2000: Climate assessment for 1999. *Bull. Amer. Meteor. Soc.*, **81** (6), S1–S50, https://doi.org/10.1175/1520-0477(2000)81[s1:CAF]2.0.CO;2.

Bian, G.-F., G.-Z. Nie, and X. Qiu, 2021: How well is outer tropical cyclone size represented in the ERA5 reanalysis dataset? *Atmos. Res.*, **249**, 105339, https://doi.org/10.1016/j.atmosres.2020.105339.

Bister, M., and K. A. Emanuel, 2002: Low frequency variability of tropical cyclone potential intensity 1. Interannual to interdecadal variability. *J. Geophys. Res.*, **107**, 4801, https://doi.org/10.1029/2001JD000776.

Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256–1271, https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2.

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

Briegleb, B. P., 1992: Delta-Eddington approximation for solar radiation in the NCAR community climate model. *J. Geophys. Res.*, **97**, 7603–7612, https://doi.org/10.1029/92JD00291.

Camargo, S. J., 2013: Global and regional aspects of tropical cyclone activity in the CMIP5 models. *J. Climate*, **26**, 9880–9902, https://doi.org/10.1175/JCLI-D-12-00549.1.

——, and A. A. Wing, 2016: Tropical cyclones in climate models. *Wiley Interdiscip. Rev.: Climate Change*, **7**, 211–237, https://doi.org/10.1002/wcc.373.

——, K. A. Emanuel, and A. H. Sobel, 2007a: Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis. *J. Climate*, **20**, 4819–4834, https://doi.org/10.1175/JCLI4282.1.

——, A. H. Sobel, A. G. Barnston, and K. A. Emanuel, 2007b: Tropical cyclone genesis potential index in climate models. *Tellus*, **59A**, 428–443, https://doi.org/10.1111/j.1600-0870.2007.00238.x.

——, M. K. Tippett, A. H. Sobel, G. A. Vecchi, and M. Zhao, 2014: Testing the performance of tropical cyclone genesis indices in future climates using the HiRAM model. *J. Climate*, **27**, 9171–9196, https://doi.org/10.1175/JCLI-D-13-00505.1.

——, A. H. Sobel, A. D. Delgenio, J. A. Jonas, M. Kelley, Y. Lu, D. A. Shaevitz, and N. Henderson, 2016: Tropical cyclones in the GISS ModelE2. *Tellus*, **68A**, 31494, https://doi.org/10.3402/tellusa.v68.31494.

——, and Coauthors, 2020: Characteristics of model tropical cyclone climatology and the large-scale environment. *J. Climate*, **33**, 4463–4487, https://doi.org/10.1175/JCLI-D-19-0500.1.

Chiodo, G., and L. Haimberger, 2010: Interannual changes in mass consistent energy budgets from ERA-Interim and satellite data. *J. Geophys. Res.*, **115**, D02112, https://doi.org/10.1029/2009JD012049.

Chou, M.-D., and M. J. Suarez, 1999: A solar radiation parameterization for atmospheric studies. Tech. Memo. NASA/TM-1999-104606, Vol. 15, 38 pp., https://gmao.gsfc.nasa.gov/pubs/docs/Chou136.pdf.

——, ——, X.-Z. Liang, and M. M.-H. Yan, 2001: A thermal infrared radiation parameterization for atmospheric studies. NASA/TM-2001-104606, Vol. 19, 54 pp., https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20010072848.pdf.

Clough, S. A., M. W. Shephard, E. J. Mlawer, J. S. Delamere, M. J. Iacono, K. Cady-Pereira, S. Boukabara, and P. D. Brown, 2005: Atmospheric radiative transfer modeling: A summary of the AER codes. *J. Quant. Spectrosc. Radiat. Transfer*, **91**, 233–244, https://doi.org/10.1016/j.jqsrt.2004.05.058.

Copernicus Climate Change Service, 2023: ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service Climate Data Store, accessed 25 March 2019, https://doi.org/10.24381/cds.adbb2d47.

Davis, C. A., 2018: Resolving tropical cyclone intensity in models. *Geophys. Res. Lett.*, **45**, 2082–2087, https://doi.org/10.1002/2017GL076966.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

DeMaria, M., 1996: The effect of vertical shear on tropical cyclone intensity change. *J. Atmos. Sci.*, **53**, 2076–2088, https://doi.org/10.1175/1520-0469(1996)053<2076:TEOVSO>2.0.CO;2.

Duvel, J.-P., S. J. Camargo, and A. H. Sobel, 2017: Role of the convection scheme in modeling initiation and intensification of tropical depressions over the North Atlantic. *Mon. Wea. Rev.*, **145**, 1495–1509, https://doi.org/10.1175/MWR-D-16-0201.1.

ECMWF, 2009: ERA-Interim Project. NCAR Computational and Information Systems Laboratory Research Data Archive, accessed 7 May 2020, https://doi.org/10.5065/D6CR5RD9.

——, 2012: ERA-Interim Project, single parameter 6-hourly surface analysis and surface forecast time series. NCAR Computational and Information Systems Laboratory Research Data Archive, accessed 7 December 2020, https://doi.org/10.5065/D64747WN.

——, 2019: ERA5 Reanalysis (0.25 degree latitude-longitude grid). NCAR Computational and Information Systems Laboratory Research Data Archive, accessed 7 December 2020, https://doi.org/10.5065/BH6N-5N20.

Elsberry, R. L., E. L. Weniger, and D. H. Meanor, 1988: A statistical tropical cyclone intensity forecast technique incorporating environmental wind and vertical wind shear information. *Mon. Wea. Rev.*, **116**, 2142–2154, https://doi.org/10.1175/1520-0493(1988)116<2142:ASTCIF>2.0.CO;2.

Emanuel, K. A., 1988: The maximum intensity of hurricanes. *J. Atmos. Sci.*, **45**, 1143–1155, https://doi.org/10.1175/1520-0469(1988)045<1143:TMIOH>2.0.CO;2.

——, 2010: Tropical cyclone activity downscaled from NOAA-CIRES reanalysis, 1908–1958. *J. Adv. Model. Earth Syst.*, **2** (1), https://doi.org/10.3894/JAMES.2010.2.1.

——, and D. S. Nolan, 2004: Tropical cyclone activity and global climate. *26th Conf. on Hurricanes and Tropical Meteorology*, Miami, FL, Amer. Meteor. Soc., 10A.2, https://ams.confex.com/ams/pdfpapers/75463.pdf.

Fouquart, Y., and B. Bonnel, 1980: Computations of solar heating of the Earth's atmosphere: A new parameterization. *Beitr. Phys. Atmos.*, **53**, 35–62.

Frank, W. M., 1982: Large-scale characteristics of tropical cyclones. *Mon. Wea. Rev.*, **110**, 572–586, https://doi.org/10.1175/1520-0493(1982)110<0572:LSCOTC>2.0.CO;2.

Freidenreich, S. M., and V. Ramaswamy, 1999: A new multiple-band solar radiative parameterization for general circulation models. *J. Geophys. Res.*, **104**, 31 389–31 409, https://doi.org/10.1029/1999JD900456.

Fujiwara, M., and Coauthors, 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417–1452, https://doi.org/10.5194/acp-17-1417-2017.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

GMAO, 2015a: MERRA-2 inst6_3d_ana_Np: 3d, 6-hourly, instantaneous, pressure-level, analysis, analyzed meteorological fields V5.12.4. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 25 March 2019, https://doi.org/10.5067/A7S6XP56VZWS.

——, 2015b: MERRA-2 tavg1_2d_flx_Nx: 2d, 1-hourly, time-averaged, single-level, assimilation, surface flux diagnostics V5.12.4. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 25 March 2019, https://doi.org/10.5067/7MCPBJ41Y0K6.

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

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

Hill, K. A., and G. M. Lackmann, 2009: Influence of environmental humidity on tropical cyclone size. *Mon. Wea. Rev.*, **137**, 3294–3315, https://doi.org/10.1175/2009MWR2679.1.

Hodges, K. I., B. J. Hoskins, J. Boyle, and C. Thorncroft, 2003: A comparison of recent reanalysis datasets using objective feature tracking: Storm tracks and tropical easterly waves. *Mon. Wea. Rev.*, **131**, 2012–2037, https://doi.org/10.1175/1520-0493(2003)131<2012:ACORRD>2.0.CO;2.

——, A. Cobb, and P. L. Vidale, 2017: How well are tropical cyclones represented in reanalysis datasets? *J. Climate*, **30**, 5243–5264, https://doi.org/10.1175/JCLI-D-16-0557.1.

JMA, 2013: JRA-55: Japanese 55-year Reanalysis, daily 3-hourly and 6-hourly data. NCAR Computational and Information Systems Laboratory Research Data Archive, accessed 22 April 2021, https://doi.org/10.5065/D6HH6H41.

——, 2020: Issue with tropical cyclone analysis in JRA-55. Japan Meteorological Agency Tech. Rep., 2 pp., https://jra.kishou.go.jp/JRA-55/document/quality_issues_20200122_en.pdf.

Jones, E., A. A. Wing, and R. Parfitt, 2021: A global perspective of tropical cyclone precipitation in reanalyses. *J. Climate*, **34**, 8461–8480, https://doi.org/10.1175/JCLI-D-20-0892.1.

Jones, S. C., 1995: The evolution of vortices in vertical shear. I: Initially barotropic vortices. *Quart. J. Roy. Meteor. Soc.*, **121**, 821–851, https://doi.org/10.1002/qj.49712152406.

Kawai, H., and T. Inoue, 2006: A simple parameterization scheme for subtropical marine stratocumulus. *SOLA*, **2**, 17–20, https://doi.org/10.2151/sola.2006-005.

Kim, D., A. H. Sobel, A. D. D. Genio, Y. Chen, S. J. Camargo, M.-S. Yao, M. Kelley, and L. Nazarenko, 2012: The tropical subseasonal variability simulated in the NASA GISS general circulation model. *J. Climate*, **25**, 4641–4659, https://doi.org/10.1175/JCLI-D-11-00447.1.

——, and Coauthors, 2018: Process-oriented diagnosis of tropical cyclones in high-resolution GCMs. *J. Climate*, **31**, 1685–1702, https://doi.org/10.1175/JCLI-D-17-0269.1.

Kim, H., M.-I. Lee, S. Kim, Y.-K. Lim, S. D. Schubert, and A. M. Molod, 2021: Representation of tropical cyclones by the Modern-Era Retrospective Analysis for Research and Applications version 2. *Asia-Pac. J. Atmos. Sci.*, **57**, 35–49, https://doi.org/10.1007/s13143-019-00169-y.

Kiranmayi, L., and E. D. Maloney, 2011: Intraseasonal moist static energy budget in reanalysis data. *J. Geophys. Res.*, **116**, D21117, https://doi.org/10.1029/2011JD016031.

Klaver, R., R. Haarsma, P. L. Vidale, and W. Hazeleger, 2020: Effective resolution in high resolution global atmospheric models for climate studies. *Atmos. Sci. Lett.*, **21**, e952, https://doi.org/10.1002/asl.952.

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

——, H. J. Diamond, J. P. Kossin, M. C. Kruk, and C. J. Schreck III, 2018: International Best Track Archive for Climate Stewardship (IBTrACS) project, version 4. NOAA National Centers for Environmental Information, accessed 25 March 2019, https://doi.org/10.25921/82ty-9e16.

Kobayashi, S., and Coauthors, 2015: The JRA-55 Reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, https://doi.org/10.2151/jmsj.2015-001.

Krouse, K. D., and A. H. Sobel, 2010: An observational study of multiple tropical cyclone events in the western North Pacific. *Tellus*, **62A**, 256–265, https://doi.org/10.1111/j.1600-0870.2009.00435.x.

Maloney, E. D., and Coauthors, 2019: Process-oriented evaluation of climate and weather forecasting models. *Bull. Amer. Meteor. Soc.*, **100**, 1665–1686, https://doi.org/10.1175/BAMS-D-18-0042.1.

Manganello, J. V., and Coauthors, 2012: Tropical cyclone climatology in a 10-km global atmospheric GCM: Toward weather-resolving climate modeling. *J. Climate*, **25**, 3867–3893, https://doi.org/10.1175/JCLI-D-11-00346.1.

McBride, J. L., and R. Zehr, 1981: Observational analysis of tropical cyclone formation. Part II: Comparison of non-developing versus developing systems. *J. Atmos. Sci.*, **38**, 1132–1151, https://doi.org/10.1175/1520-0469(1981)038<1132:OAOTCF>2.0.CO;2.

McCarty, W., L. Coy, R. Gelaro, A. Huang, D. Merkova, E. B. Smith, M. Sienkiewicz, and K. Wargan, 2016: MERRA-2 input observations: Summary and assessment. NASA Tech. Rep. NASA/TM-2016-104606, Vol. 46, 51 pp., https://gmao.gsfc.nasa.gov/pubs/docs/McCarty885.pdf.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogenous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, https://doi.org/10.1029/97JD00237.

Molod, A., L. Takacs, M. Suarez, and J. Bacmeister, 2015: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geosci. Model Dev.*, **8**, 1339–1356, https://doi.org/10.5194/gmd-8-1339-2015.

Moon, Y., D. Kim, S. J. Camargo, A. A. Wing, K. A. Reed, M. F. Wehner, and M. M. Zhao, 2020a: A new method to construct a horizontal resolution-dependent wind speed adjustment

factor for tropical cyclones in global climate model simulations. *Geophys. Res. Lett.*, **47**, e2020GL087528, https://doi.org/10.1029/2020GL087528.

——, and Coauthors, 2020b: Azimuthally averaged wind and thermodynamic structures of tropical cyclones in global climate models and their sensitivity to horizontal resolution. *J. Climate*, **33**, 1575–1595, https://doi.org/10.1175/JCLI-D-19-0172.1.

——, and Coauthors, 2022: An evaluation of tropical cyclone rainfall structures in the HighResMIP simulations against satellite observations. *J. Climate*, **35**, 7315–7338, https://doi.org/10.1175/JCLI-D-21-0564.1.

Moorthi, S., and M. J. Suarez, 1992: Relaxed Arakawa-Schubert: A parameterization of moist convection for general circulation models. *Mon. Wea. Rev.*, **120**, 978–1002, https://doi.org/10.1175/1520-0493(1992)120<0978:RASAPO>2.0.CO;2.

——, H.-L. Pan, and P. Caplan, 2001: Changes to the 2001 NCEP Operational MRV/AVN Global Analysis/Forecast System. NWS Tech. Procedures Bull. 484, 14 pp.

Murakami, H., 2014: Tropical cyclones in reanalysis data sets. *Geophys. Res. Lett.*, **41**, 2133–2141, https://doi.org/10.1002/2014GL059519.

——, and Coauthors, 2012: Future changes in tropical cyclone activity projected by the new high-resolution MRI-AGCM. *J. Climate*, **25**, 3237–3260, https://doi.org/10.1175/JCLI-D-11-00415.1.

——, and Coauthors, 2015: Simulation and prediction of category 4 and 5 hurricanes in the high-resolution GFDL HiFLOR coupled climate models. *J. Climate*, **28**, 9058–9079, https://doi.org/10.1175/JCLI-D-15-0216.1.

Parker, W. S., 2016: Reanalyses and observations: What's the difference? *Bull. Amer. Meteor. Soc.*, **97**, 1565–1572, https://doi.org/10.1175/BAMS-D-14-00226.1.

Reed, K. A., and C. Jablonowski, 2011: Impact of physical parameterizations on idealized tropical cyclones in the Community Atmosphere Model. *Geophys. Res. Lett.*, **38**, L04805, https://doi.org/10.1029/2010GL046297.

——, J. T. Bacmeister, N. A. Rosenbloom, M. F. Wehner, S. C. Bates, P. H. Lauritzen, J. E. Truesdale, and C. Hannay, 2015: Impact of the dynamical core on the direct simulation of tropical cyclones in a high-resolution global model. *Geophys. Res. Lett.*, **42**, 3603–3608, https://doi.org/10.1002/2015GL063974.

Rienecker, M. M., and Coauthors, 2008: The GEOS-5 Data Assimilation System—Documentation of versions 5.0.1, 5.1.0, and 5.2.0. NASA/TM–2008–104606, Vol. 27, 118 pp., https://gmao.gsfc.nasa.gov/pubs/docs/GEOS5_104606-Vol27.pdf.

Roberts, M. J., and Coauthors, 2015: Tropical cyclones in the UPSCALE ensemble of high-resolution global climate models. *J. Climate*, **28**, 574–596, https://doi.org/10.1175/JCLI-D-14-00131.1.

——, and Coauthors, 2020: Impact of model resolution on tropical cyclone simulation using the HighResMIP–PRIMAVERA multimodel ensemble. *J. Climate*, **33**, 2557–2583, https://doi.org/10.1175/JCLI-D-19-0639.1.

Ruppert, J. H., Jr., A. A. Wing, X. Tang, and E. L. Duran, 2020: The critical role of cloud–infrared radiation feedback in tropical cyclone development. *Proc. Natl. Acad. Sci. USA*, **117**, 27 884–27 892, https://doi.org/10.1073/pnas.2013584117.

Russotto, R. D., and Coauthors, 2022: Improved representation of tropical cyclones in the NASA GISS-E3 GCM. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002601, https://doi.org/10.1029/2021MS002601.

Saha, S., and Coauthors, 2010a: NCEP Climate Forecast System Reanalysis (CFSR) 6-hourly products, January 1979 to December 2010. NCAR Computational and Information Systems Laboratory Research Data Archive, accessed 26 October 2020, https://doi.org/10.5065/D69K487J.

——, and Coauthors, 2010b: NCEP Climate Forecast System Reanalysis (CFSR) selected hourly time-series products, January 1979 to December 2010. NCAR Computational and Information Systems Laboratory Research Data Archive, accessed 25 March 2019, https://doi.org/10.5065/D6513W89.

——, and Coauthors, 2010c: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1.

——, and Coauthors, 2011: NCEP Climate Forecast System Version 2 (CFSv2) 6-hourly Products. NCAR Computational and Information Systems Laboratory Research Data Archive, accessed 26 October 2020, https://doi.org/10.5065/D61C1TXF.

——, and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1.

Schenkel, B. A., and R. E. Hart, 2012: An examination of tropical cyclone position, intensity, and intensity life cycle within atmospheric reanalysis datasets. *J. Climate*, **25**, 3453–3475, https://doi.org/10.1175/2011JCLI4208.1.

——, N. Lin, D. Chavas, M. Oppenheimer, and A. Brammer, 2017: Evaluating outer tropical cyclone size in reanalysis datasets using QuikSCAT data. *J. Climate*, **30**, 8745–8762, https://doi.org/10.1175/JCLI-D-17-0122.1.

Shaevitz, D. A., and Coauthors, 2014: Characteristics of tropical cyclones in high-resolution models in the present climate. *J. Adv. Model. Earth Syst.*, **6**, 1154–1172, https://doi.org/10.1002/2014MS000372.

Simmons, A., and Coauthors, 2020: Global stratospheric temperature bias and other stratospheric aspects of ERA5 and ERA5.1. ECMWF Tech. Memo. 859, 40 pp., https://www.ecmwf.int/sites/default/files/elibrary/2020/19362-global-stratospheric-temperature-bias-and-other-stratospheric-aspects-era5-and-era51.pdf.

Slocum, C. J., M. N. Razin, J. A. Knaff, and J. P. Stow, 2022: Does ERA5 mark a new era for resolving the tropical cyclone environment? *J. Climate*, **35**, 7147–7164, https://doi.org/10.1175/JCLI-D-22-0127.1.

Sobel, A., S. Wang, and D. Kim, 2014: Moist static energy budget of the MJO during DYNAMO. *J. Atmos. Sci.*, **71**, 4276–4291, https://doi.org/10.1175/JAS-D-14-0052.1.

Thépaut, J.-N., P. Courtier, G. Belaud, and G. Lematre, 1996: Dynamical structure functions in a four-dimensional variational assimilation: A case study. *Quart. J. Roy. Meteor. Soc.*, **122**, 535–561, https://doi.org/10.1002/qj.49712253012.

Tiedtke, M., 1983: The sensitivity of the time-mean large-scale flow to cumulus convection in the ECMWF model. *Workshop on Convection in Large-Scale Numerical Models*, Reading, UK, European Centre for Medium-Range Weather Forecasts, 297–316, https://www.ecmwf.int/node/12733.

——, 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2.

——, 1993: Representation of clouds in large-scale models. *Mon. Wea. Rev.*, **121**, 3040–3061, https://doi.org/10.1175/1520-0493(1993)121%3C3040:ROCILS%3E2.0.CO;2.

Tippett, M. K., S. J. Camargo, and A. H. Sobel, 2011: A Poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis. *J. Climate*, **24**, 2335–2357, https://doi.org/10.1175/2010JCLI3811.1.

Ullrich, P. A., and C. M. Zarzycki, 2017: TempestExtremes: A framework for scale-insensitive pointwise feature tracking on

unstructured grids. *Geosci. Model Dev.*, **10**, 1069–1090, https://doi.org/10.5194/gmd-10-1069-2017.

Vannière, B., and Coauthors, 2020: The moisture budget of tropical cyclones in HighResMIP models: Large-scale environmental balance and sensitivity to horizontal resolution. *J. Climate*, **33**, 8457–8474, https://doi.org/10.1175/JCLI-D-19-0999.1.

Vidale, P. L., and Coauthors, 2021: Impact of stochastic physics and model resolution on the simulation of tropical cyclones in climate GCMs. *J. Climate*, **34**, 4315–4341, https://doi.org/10.1175/JCLI-D-20-0507.1.

Walsh, K. J. E., M. Fiorino, C. W. Landsea, and K. L. McInnes, 2007: Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses. *J. Climate*, **20**, 2307–2314, https://doi.org/10.1175/JCLI4074.1.

——, and Coauthors, 2016: Tropical cyclones and climate change. *Wiley Interdiscip. Rev.: Climate Change*, **7**, 65–89, https://doi.org/10.1002/wcc.371.

Wang, B., and H. Murakami, 2020: Dynamic genesis potential index for diagnosing present-day and future global tropical cyclone genesis. *Environ. Res. Lett.*, **15**, 114008, https://doi.org/10.1088/1748-9326/abbb01.

Wehner, M. F., and Coauthors, 2014: The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1. *J. Adv. Model. Earth Syst.*, **6**, 980–997, https://doi.org/10.1002/2013MS000276.

Wing, A. A., 2022: Acceleration of tropical cyclone development by cloud-radiative feedbacks. *J. Atmos. Sci.*, **79**, 2285–2305, https://doi.org/10.1175/JAS-D-21-0227.1.

——, and K. A. Emanuel, 2014: Physical mechanisms controlling self-aggregation of convection in idealized numerical modeling simulations. *J. Adv. Model. Earth Syst.*, **6**, 59–74, https://doi.org/10.1002/2013MS000269.

——, and T. W. Cronin, 2016: Self-aggregation of convection in long channel geometry. *Quart. J. Roy. Meteor. Soc.*, **142** (694), 1–15, https://doi.org/10.1002/qj.2628.

——, and Coauthors, 2019: Moist static energy budget analysis of tropical cyclone intensification in high-resolution climate models. *J. Climate*, **32**, 6071–6095, https://doi.org/10.1175/JCLI-D-18-0599.1.

Wu, S.-N., B. J. Soden, and D. S. Nolan, 2021: Examining the role of cloud radiative interactions in tropical cyclone development using satellite measurements and WRF simulations. *Geophys. Res. Lett.*, **48**, e2021GL093259, https://doi.org/10.1029/2021GL093259.

——, ——, and G. J. Alaka Jr., 2023: The influence of radiation on the prediction of tropical cyclone intensification in a forecast model. *Geophys. Res. Lett.*, **50**, e2022GL099442, https://doi.org/10.1029/2022GL099442.

Xie, S., and M. Zhang, 2000: Impact of the convection triggering function on single-column model simulations. *J. Geophys. Res.*, **105**, 14 983–14 996, https://doi.org/10.1029/2000JD900170.

Xu, K.-M., and D. A. Randall, 1996: A semiemperical cloudiness parameterization for use in climate models. *J. Atmos. Sci.*, **53**, 3084–3102, https://doi.org/10.1175/1520-0469(1996)053<3084:ASCPFU>2.0.CO;2.

Zarzycki, C. M., D. R. Thatcher, and C. Jablonowski, 2017: Objective tropical cyclone extratropical transition detection in high-resolution reanalysis and climate model data. *J. Adv. Model. Earth Syst.*, **9**, 130–148, https://doi.org/10.1002/2016MS000775.

——, P. A. Ullrich, and K. A. Reed, 2021: Metrics for evaluating tropical cyclones in climate data. *J. Appl. Meteor. Climatol.*, **60**, 643–660, https://doi.org/10.1175/JAMC-D-20-0149.1.

Zhang, B., B. J. Soden, G. A. Vecchi, and W. Yang, 2021: The role of radiative interactions in tropical cyclone development under realistic boundary conditions. *J. Climate*, **34**, 2079–2091, https://doi.org/10.1175/JCLI-D-20-0574.1.

Zhang, W., and Coauthors, 2021: Tropical cyclone precipitation in the HighResMIP atmosphere-only experiments of the PRIMAVERA project. *Climate Dyn.*, **57**, 253–273, https://doi.org/10.1007/s00382-021-05707-x.

Zhao, M., I. M. Held, and S.-J. Lin, 2012: Some counterintuitive dependencies of tropical cyclone frequency on parameters in a GCM. *J. Atmos. Sci.*, **69**, 2272–2283, https://doi.org/10.1175/JAS-D-11-0238.1.

Zhao, Q., and F. H. Carr, 1997: A prognostic cloud scheme for operational NWP models. *Mon. Wea. Rev.*, **125**, 1931–1953, https://doi.org/10.1175/1520-0493(1997)125<1931:APCSFO>2.0.CO;2.