Check for updates

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2020MS002447

Evaluating Precipitation Errors Using the Environmentally Conditioned Intensity-Frequency Decomposition Method

A. Di Luca[1,2,3], D. Argüeso[4], S. Sherwood[2,3], and J. P. Evans[2,3]

[1]Département des sciences de la Terre et de l'atmosphère, Université du Québec à Montréal, Montreal, QC, Canada, [2]Climate Change Research Centre, University of New South Wales, Sydney, NSW, Australia, [3]ARC Centre of Excellence for Climate Extremes, Sydney, NSW, Australia, [4]Physics Department, University of Balearic Islands, Palma, Spain

Key Points:

- Multiple high-resolution simulations produce precipitation too often with underestimated intensity leading to strong error compensation
- Errors in precipitation are quantified using novel metrics that prevent error compensation showing value compared with traditional metrics
- Convection permitting simulations outperform the representation of precipitation compared to simulations using parameterized convection

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Di Luca,
di_luca.alejandro@uqam.ca

Citation:

Di Luca, A., Argüeso, D., Sherwood, S., & Evans, J. P. (2021). Evaluating precipitation errors using the environmentally conditioned intensity-frequency decomposition method. *Journal of Advances in Modeling Earth Systems*, *13*, e2020MS002447. https://doi.org/10.1029/2020MS002447

Received 28 DEC 2020
Accepted 30 JUN 2021

© 2021. The Authors. Journal of Advances in Modeling Earth Systems published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Abstract A fundamental issue when evaluating the simulation of precipitation is the difficulty of quantifying specific sources of errors and recognizing compensation of errors. We assess how well a large ensemble of high-resolution simulations represents the precipitation associated with strong cyclones. We propose a framework to breakdown precipitation errors according to different dynamical (vertical velocity) and thermodynamical (vertically integrated water vapor) regimes and the frequency and intensity of precipitation. This approach approximates the error in the total precipitation of each regime as the sum of three terms describing errors in the large-scale environmental conditions, the frequency of precipitation and its intensity. We show that simulations produce precipitation too often, that its intensity is too weak, that errors are larger for weak than for strong dynamical forcing and that biases in the vertically integrated water vapor can be large. Using the error breakdown presented above, we define four new error metrics differing on the degree to which they include the compensation of errors. We show that convection-permitting simulations consistently improve the simulation of precipitation compared to coarser-resolution simulations using parameterized convection, and that these improvements are revealed by our new approach but not by traditional metrics which can be affected by compensating errors. These results suggest that convection-permitting models are more likely to produce better results for the right reasons. We conclude that the novel decomposition and error metrics presented in this study give a useful framework that provides physical insights about the sources of errors and a reliable quantification of errors.

Plain Language Summary The simulations of complex physical processes always entail various sources of errors. These errors can be of different sign and can consequently cancel each other out when using traditional performance metrics such as the bias error metric. We present a formal framework that allows us to approximate precipitation according to three terms that describe different aspects of the rainfall field including large-scale environmental conditions and the frequency and intensity of rainfall. We apply the methodology to a large ensemble of high-resolution simulations representing the precipitation associated with strong cyclones in eastern Australia. We show that simulations produce precipitation too often, with an intensity that is too weak leading to strong compensation. We further define new error metrics that explicitly quantify the degree of error compensation when simulating precipitation. We show that convection-permitting simulations consistently improve the performance compared to coarser resolution simulations using parameterized convection and that these improvements are only revealed when using the new error metrics but are not apparent in traditional metrics (e.g., bias).

1. Introduction

The simulation of precipitation constitutes a great challenge for current climate models because of the wide range of physical processes and interactions that are involved (e.g., Stensrud, 2007). First, a climate model needs to simulate well the large-scale environmental conditions that favor the occurrence of rainfall events including various types of hydrodynamic instabilities (convective and baroclinic instability) and the resulting phenomena such as midlatitude cyclones, frontal structures, and shallow and deep convection (Catto et al., 2010; Di Luca, Argüeso, et al., 2016; Di Luca, Evans, et al., 2016; Govekar et al., 2014; Zappa et al., 2013). Once the precipitation is triggered, the model needs to simulate well a range of features such

as the convergence of moisture, which depends on moisture fluxes within the boundary layer, and the overall precipitation efficiency, which depends on microphysical processes (Bao & Sherwood, 2019; Singh & O'Gorman, 2014). Finally, the overall duration of the precipitating event will be determined by the evolution of the large-scale conditions and the resulting feedbacks from the precipitation process itself, including changes in the vertical profile of temperature and water vapor (Stensrud, 2007).

A direct consequence of having so many distinct processes influencing precipitation is that errors can arise from a variety of sources. Consequently, models can sometimes produce the right results for the wrong reasons by compensating errors from different processes having opposite sign (e.g., Di Luca et al., 2020; Eyring et al., 2016; Palmer, 2016). To better understand the sources of errors, studies sometimes separate the rainfall field according to different terms such as the number and intensity of precipitating events (e.g., Dai et al., 2017; Rajczak & Schär, 2017; Williams & Tselioudis, 2007), the type of phenomenon behind rainfall events (e.g., frontal or low pressure systems; Catto et al., 2015; Zappa et al., 2015) or the type of weather regime affecting the region (e.g., Brown et al., 2010). For example, using an objective algorithm to identify fronts, Catto et al. (2015) separated the error in total precipitation into frontal and nonfrontal regimes, each of them further separated according to their frequency of occurrence and their mean intensity. They found that while the frequency of fronts was well represented by models, the frequency of precipitation associated with fronts was too high and their intensity too low, thus leading to relatively small errors due to error compensation.

In addition, the ability of climate models to simulate different characteristics of precipitation at subdaily and daily time scales has been shown to depend on their horizontal resolution (Ban et al., 2014; Di Luca, Argüeso, et al., 2016; Di Luca, Evans, et al., 2016; Knist et al., 2018; Prein et al., 2016; Rajczak & Schär, 2017). An important issue sometimes overlooked when assessing aspects such as the frequency and intensity of precipitation across data sets with different resolutions is related with whether we evaluate how realistic the model is or how good the model is (e.g., C.-T. Chen & Knutson, 2008). In the first case, we will possibly use the highest resolution observations available (e.g., rain gauges stations) and will directly compare observations and simulations to try to answer which one is closer to "real" values. In the second case, we will possibly use gridded observations that represent the same temporal and spatial scales as our climate models, and we will try to answer how well the model reproduces the quantity that it was designed for (the gridbox-mean rainfall). Prein et al. (2016) evaluated precipitation statistics over Europe in an ensemble of Regional Climate Model (RCM) simulations with 0.11° and 0.44° grid spacings. Their comparison, performed at the common 0.44° scale, showed that higher resolution models were able to reduce mean biases and improve spatial patterns of seasonal and daily extreme rainfall. They attributed these improvements to the better representation of orography and to capturing larger scales in convection by the resolved-scale dynamics during summer. Knist et al. (2018) evaluated the ability of Weather Research and Forecasting (WRF) model to simulate subdaily precipitation characteristics over a domain covering Germany and Switzerland using a double nesting setup with 3 and 12-km grid spacings. At hourly time scales, they found that both WRF resolutions produced too many precipitating events and underestimated the intensity of very heavy events compared with rain gauge estimations during boreal winter. During boreal summer, precipitation statistics derived from the 3-km version were much closer to observations than those from the 12-km version, showing the importance of horizontal resolution in convective environments. Ban et al. (2014) also evaluated hourly rainfall frequency and intensity in two versions of a RCM with horizontal grid spacings of 2 and 12 km. They found that the 2-km grid spacing RCM was able to better represent rainfall statistics compared to the 12-km version even after the 2-km model was upscaled over the 12-km grid mesh.

Midlatitude cyclones are responsible for much of the high-impact weather affecting the east coast of Australia. East coast lows (ECLs), as they are known in Australia, often grow and develop from baroclinic instability (i.e., pure extratropical cyclones) but can also include extratropical transitions of tropical cyclones and hybrid cyclones, which are strongly influenced by diabatic processes including exchanges of heat and moisture with the underlying Tasman sea (Cavicchia et al., 2020; Dowdy et al., 2019; Quinting et al., 2019). ECLs can produce very heavy rainfall over periods ranging from minutes to days. Deep convection and thunderstorms dominate heavy rainfall rates at subdaily time scales and synoptic scale organization, including cold and warm fronts, generally governs large accumulations over multi-day periods (e.g., Catto & Pfahl, 2013; Dowdy & Catto, 2017). For instance, Woolgoolga near the northern New South Wales (NSW)

coast recorded 113 mm in 1 h during the passage of the June 2016 ECL (Louis et al., 2016) while Mangrove Mountain near the central NSW coast recorded over 400 mm in 72 h during the passage of the Pasha Bulker ECL in June 2007 (Mills et al., 2010).

In this study, we develop a new methodology to evaluate the ability of an ensemble of high-resolution simulations to represent the rainfall associated with some of the most extreme midlatitude cyclones that affected the east coast of Australia over recent decades. We focus on the ability of simulations to reproduce the sensitivity of multiple rainfall characteristics to variations in dynamical and thermodynamical environmental conditions. This is done by simultaneously applying two independent but complementary decompositions of the rainfall field. First, the total precipitation is separated into different dynamical and thermodynamical regimes associated with specific large-scale conditions (e.g., Bony et al., 2004; Brown et al., 2010; Holloway & Neelin, 2009; Sahany et al., 2014). Second, for each regime, the total precipitation is separated into the product of the frequency of precipitating events and their mean intensity (e.g., Dai et al., 2017). Combining both decompositions, we show that the error in total precipitation can be approximated according to three terms representing errors in different aspects of the precipitation process: (a) the large-scale environmental regimes, (b) the frequency of rainfall, and (c) the amount of rainfall (i.e., intensity).

The decomposition is applied to observations and an ensemble of regional climate model simulations of 11 storms that affected the Sydney area between 2001 and 2016. The ensemble is constructed using a single regional climate model but includes simulations using multiple resolutions (2, 8, and 24 km grid spacings) and multiple subgrid-scale parameterizations allowing the assessment of the models' sensitivity to the specific configuration. The evaluation is performed using several state-of-the-art high-resolution observational products merging in situ and satellite data sets.

The study is organized as follows. In Section 2, we describe the ensemble of simulations and the reference data sets used to assess the simulations. In Section 3, we introduce the methodology to decompose precipitation errors in various terms and we define different error metrics to quantify the degree of error compensation in simulations. In Section 4, we apply the decomposition method to assess errors in the ensemble of simulations and discuss in detail the decomposition of errors over ocean/land grid points including the influence of resolution, physics and other aspects of the ensemble. Some discussion about the observational uncertainty and other issues is presented in Section 5, and a summary and conclusions are given in Section 6.

2. Data

2.1. Observed Data

Five high-spatial resolution (≤30 km) land and ocean satellite-based precipitation products are considered in this study (Table 1). All products include estimates from passive microwave and infrared data from multiple satellites, taking advantage of the great accuracy of passive microwave estimates and the high-temporal frequency of infrared estimates (e.g., Kidd & Levizzani, 2011). All products are corrected using land gauge precipitation analysis although the specific type of correction applied (e.g., temporal resolution of the input) depends on the product. Tropical Rainfall Measuring Mission (TRMM)-3B42 and Integrated Multi-satellitE Retrievals for GPM (IMERG) precipitation fields are corrected using the Global Precipitation Climatology Center (GPCC) monthly gauge precipitation analyses (Huffman et al., 2007). Climate Prediction Center MORPHing (CMORPH)-CRT satellite estimations are corrected using a probability density function matching technique against the National Oceanic and Atmospheric Administration Climate Prediction Center's (CPC) daily gauge analysis. Global Satellite Mapping of Precipitation (GSMaP) is also corrected using CPC daily gauges. The 3-hourly temporal variability of the Multi-Source Weighted-Ensemble Precipitation (MSWEP) product was determined by weighted averaging of precipitation anomalies from seven data sets: two based on gridded in situ observations (CPC and GPCC), three on satellite remote sensing (CMORPH, GSMaP, and TRMM-3B42), and two reanalyzes (ERA-Interim and JRA-55).

Two additional daily rainfall data sets are used to evaluate mean precipitation over land: the Australian Gridded Climate Data (AGCD) and the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) data sets (see Table 1). Both data sets are available at high-spatial resolution (0.05°) but relatively low temporal resolution (24 h) over land grid points so they are only used to assess precipitation mean fields.

Table 1
Native Temporal and Spatial Resolution, Coverage (L: Land; O: Ocean) and Input Data Sources of the Observation-Based Data Sets Used in the Study

Abb.	Name/Version	Δt (h)	Δx (°)	Cover.	Data source	Reference
CMORPH-CRT	CPC MORPHing technique bias corrected V1.0	0.5	0.25	L/O	S/G	Joyce et al. (2004); Xie et al. (2017); Xie et al. (2019)
GSMaP	Global Satellite Mapping of Precipitation V6 (GNRT6_V1.1)	1	0.1	L/O	S/G	Kubota et al. (2007); Kubota et al. (2020); Mega et al. (2019)
IMERG	Integrated Multi-satellitE Retrievals for GPM V6	1	0.1	L/O	S/G	Huffman, Bolvin, et al. (2019); Huffman, Stocker, et al. (2019)
MSWEP	Multi-Source Weighted-Ensemble Precipitation V2.2	3	0.1	L/O	S/G/R	Beck et al. (2017); Beck et al. (2019)
TRMM-3B42	Tropical Rainfall Measuring Mission V7	3	0.25	L/O	S/G	Huffman et al. (2007); TRMM (2011)
AGCD	Australian Gridded Climate Data	24	0.05	L	G	Jones et al. (2009)
CHIRPS	Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) V2.0	24	0.05	L	S/G	Funk et al. (2015)

Data sources are satellite (S), gauges (G), and reanalysis (R).

Observed large-scale environmental conditions are estimated using the two latest reanalyzes from the European Centre for Medium-Range Weather Forecasts (ECMWF): ERA-Interim and ERA5. Over the region of interest, the ERA-Interim (ERAI) reanalysis (Dee et al., 2011) has been shown to perform best regarding the number and intensity of ECLs when compared to other recent reanalyzes such as MERRA, CFSR, and JRA55 (Pepler et al., 2017). The ERA5 reanalysis (Hersbach et al., 2020) is the newest reanalysis product by ECMWF and improves upon ERAI in several ways including increasing the horizontal resolution and developments in the model physics, numerics, and data assimilation scheme.

2.2. Simulated Data

We use a subensemble of the High-Resolution Modeling of Extreme Storms over the East Coast of Australia data set (Di Luca et al., 2021) where version 3.6 of the WRF model (Powers et al., 2017; W. Skamarock et al., 2008) was used to perform the simulations. WRF solves the compressible, non-hydrostatic Euler equations using a terrain-following vertical coordinate. We use the Advanced Research WRF dynamical solver and a triple nesting set up with a focus over eastern Australia (see Figure 1). The outer domain corresponds to the Coordinated Regional Climate Downscaling Experiment (CORDEX) Australasia domain and is discretized using a 24-km horizontal grid spacing (i.e., WRF24, 289 × 431). WRF24 is initialized and driven at the lateral boundaries of the domain using 6-hourly data from the ECMWF ERAI reanalysis (Dee et al., 2011). In addition, the geopotential height and wind components above the planetary boundary layer are nudged toward ERAI using a spectral nudging technique for wavelengths ≥600 km (W. Skamarock et al., 2008). This nudging ensures that the solution inside the large CORDEX domain does not diverge substantially from the driving reanalysis, which facilitates the comparison between simulations and observations (e.g., Giorgi & Bi, 2000; Separovic et al., 2011). WRF24 is used to drive, at every model time step using one-way nesting, an 8-km grid spacing configuration (WRF8; 405 × 435) which in turn is used to drive a 2-km grid spacing configuration (WRF2; 960 × 1,080) over domains centered on eastern Australia (see Figure 1). WRF8 and WRF2 are driven at the lateral boundaries only, without any spectral nudging. All simulations were performed using 30 sigma levels in the vertical and the Unified Noah land surface model (F. Chen & Dudhia, 2001). Times steps are 120, 40, and 10 s for WRF24, WRF8, and WRF2 simulations, respectively. The computational cost of running the WRF2 configuration is about 100 times more expensive than running WRF24.

We explore the role played by internal variability and by structural model uncertainties in simulations by performing all experiments using a five-member multi-physics ensemble. The multi-physics ensemble is

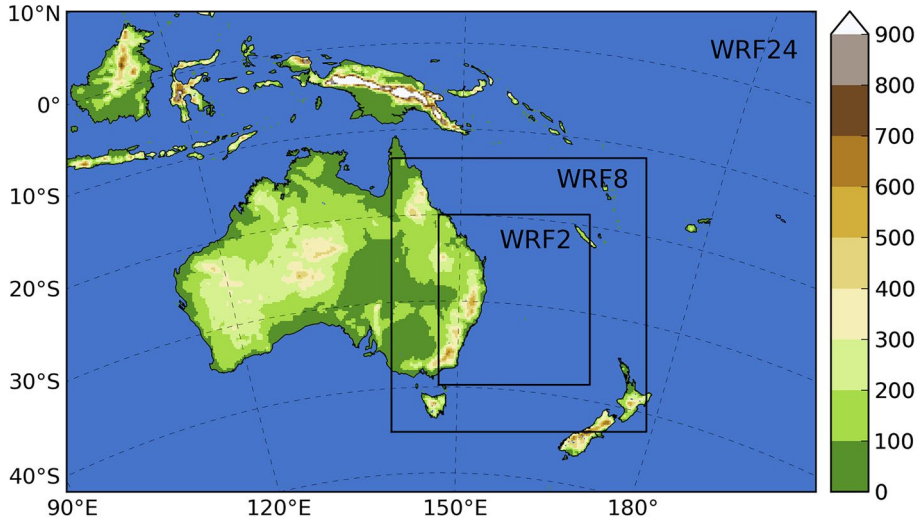

Figure 1. Domains used to run the Weather Research and Forecasting (WRF) model with grid spacings of 24 (WRF24), 8 (WRF8), and 2 km (WRF2). The region of analysis corresponds to the WRF2 domain and contains a total of 1,510 land grid points and 5,770 ocean grid points at the common 24-km resolution.

constructed around a control (CTL) run by using different formulations of subgrid-scale process. Table 2 shows details on the configuration of the CTL run and the four perturbed physics members that include modifications in the cumulus (CU), the surface and planetary boundary layer (PBL), the radiation (RAD), and the microphysics (MPS) schemes. Members are denoted according to the physical scheme that is being changed compared to the CTL run. For example, the cumulus member (denoted as CU) uses the same subgrid-scale schemes as the CTL run but a different cumulus scheme (e.g., Kain-Fritsch scheme instead of the Betts-Miller-Jancic scheme). Subgrid-scale convective processes are represented using parameterization schemes in WRF24 and WRF8. WRF2 simulations are ran without parameterized convection and relies on the explicit representation of convective processes.

Table 2
Subgrid-Scale Physics Schemes Used in Each of the Five Members of the Multi-Physics Ensemble

	CTL	CU	PBL	RAD	MPS
Cumulus	BMJ	**KF**	BMJ	BMJ	BMJ
Surface/Planetary Boundary Layer	YSU	YSU	**MYJ**	YSU	YSU
Longwave	RRTM	RRTM	RRTM	CAM	RRTM
Shortwave	Dudhia	Dudhia	Dudhia	**CAM**	Dudhia
Microphysics	WSM6	WSM6	WSM6	WSM6	**Thompson**

Note. Members are denoted according to the physical scheme that is being changed compared to the control (CTL) run. For example, the CU member is run using the same parameterizations as the CTL member with the exception of the cumulus scheme that is replaced by KF. MYJ (Janjic, 1994); YSU (Hong et al., 2006); KF (Kain & Fritsch, 1993; Kain, 2004); BMJ (Betts, 1986; Janjic, 1994, 2000); WSM6 (Lim & Hong, 2010); Dudhia (Dudhia, 1989), RRTM (Mlawer et al., 1997); CAM (Collins et al., 2004). For more details about parameterizations see W. Skamarock et al. (2008). The cumulus scheme is only used in WRF24 and WRF8 while WRF2 is run without cumulus parameterization. Subgrid-Scale physic schemes in bold denote those that are changed compared to the control member (CTL).

We selected a total of 11 events featuring a strong ECL over the eastern coast of Australia. Events were selected partly based on their impacts around the Sydney area by looking at the accumulated and maximum 3-hourly precipitation rates and 3-hourly wind speeds from ERAI close to Sydney. However, we also included some of the most iconic ECL events in recent times such as the three systems in June 2007 (including the "Pasha Bulker") and the event in June 2016. All events were simulated for a total of 8 days, starting about 4 days before the storm peaked near the Sydney's area. Table 3 provides the number and initial date of all simulations and includes the total precipitation and the highest 3-hourly rate observed around Sydney for each event. Figures S1 and S2 show a snapshot of each selected storm about 4 days (99 h) after the beginning of each simulation. Snapshots show the mean sea level pressure field from the ERA5 reanalysis together with the 3-hourly rainfall from the IMERG data set. The storms show a variety of mean sea level pressure patterns with some systems showing highly asymmetric rain and wind fields (e.g., #1) and others showing somewhat symmetric fields (e.g., #11), suggesting a mixed of extratropical and hybrid cyclones (Cavicchia et al., 2020; Dowdy et al., 2019; Quinting et al., 2019). Note that most events occurred during the southern hemisphere winter months (six events) and one event in summer, two in fall and two in spring months.

Two sets of simulations were performed for all events using different sources of oceanic boundary conditions (OBCs). One set of simulations

Table 3
All Events Were Simulated for 8 Days Starting at 00:00 on the Initial Date

#	Initial date	Total	Max. rate
1	2007-06-04	155.3	59.7
2	2007-06-13	200.3	42.6
3	2007-06-22	39.0	18.2
4	2001-07-23	86.3	27.4
5	2005-03-18	39.4	23.4
6	2008-09-02	67.6	21.2
7	2015-04-18	213.6	40.8
8	2008-08-18	20.7	16.9
9	2013-02-17	74.6	56.7
10	2016-06-01	168.3	36.2
11	2006-09-03	111.6	45.8

Note. Total (in mm) and maximum 3-hourly precipitation rate (in mm $(3\ h)^{-1}$) averaged within a 100-km radius around Sydney are provided in columns 2 and 3. Precipitation rates were estimated using the median across five satellite-observed data sets (IMERG, CMORPH-CRT, GSMaP, TRMM-3B42 and MSWEP).

Abbreviations: CMORPH-CRT, CPC MORPHing technique bias corrected; GSMaP, Global Satellite Mapping of Precipitation; IMERG, Integrated Multi-satellitE Retrievals for GPM; MSWEP, Multi-Source Weighted-Ensemble Precipitation; TRMM, Tropical Rainfall Measuring Mission.

was performed using sea surface temperatures (SSTs) directly obtained from the ERAI reanalysis (~0.75°). A second set of simulations was performed using the higher spatial resolution SSTs (~0.1°) from the Bluelink ReANalysis (BRAN) data set (Oke et al., 2013). BRAN uses an eddy resolving ocean model, constrained with Argo, SST, and altimeter observations through a data assimilation system. Over eastern Australia, Chambers et al. (2014, 2015) have shown that fine-scale SST gradients play a crucial role in determining the location and intensity of ECL-related rainfall. The comparison between both sets of simulations allows assessment of the influence of high-resolution SST gradients in the simulation of the storms.

The full ensemble thus comprises a total of 330 simulations that can be separated as follows:

$$E = S_{r,p,e,b} \quad (1)$$

where subindices r, p, e, and b denote the various resolutions (3), physics (5), events (11), and OBCs (2), respectively.

3. Methods

3.1. Regridding Precipitation

Some characteristics of precipitation such as the frequency of dry periods or the intensity of the most extreme rates strongly depend on the spatial and temporal resolution of the data. Consequently, when directly comparing precipitation rates from data sets with different temporal and/or spatial resolutions, it is generally not possible to determine whether differences are attributable to the quality of each data set or the spatiotemporal scales they actually represent (e.g., C.-T. Chen & Knutson, 2008).

The issue of comparing data sets with different resolutions is commonly approached by postprocessing (remapping) observed and simulated data so that the resulting data provide information at similar spatiotemporal scales (e.g., Di Luca, Argüeso, et al., 2016; Di Luca, Evans, et al., 2016; Prein et al., 2016). In this study, we perform the comparison between simulations and observations using a common, coarse resolution, grid mesh of about 24-km in latitude and longitude taken from the WRF24 grid mesh. High-resolution simulations (e.g., WRF2 and WRF8) and observations (e.g., IMERG) are upscaled into the 24-km grid using a conservative mapping technique. The remapping ensures that the information at spatial scales finer than 24 km is filtered out although such a methodology does not alter the influence that fine scales could have had on the large scales. Similarly, the comparison is performed using 3-hourly precipitation fields as this is the highest temporal resolution that is common across all observations and models.

Figure 2 illustrates this remapping by showing the 3-hourly precipitation field at the native resolution (left panels) and the resulting field remapped at the common 24-km grid mesh. Results are shown for the CTL WRF2 simulation and the IMERG observations for a single 3-hourly period. The domain-average and the domain-maximum precipitation rates (see bottom-left corners of each plot) show that the domain-mean values remain largely unchanged but extreme rates can decrease substantially after the remapping is performed. The smoothing of extremes is stronger for the WRF2 simulation compared with the IMERG product because the jump in resolution between the native and the common grid is much larger for WRF2 (a factor of 12 for WRF2 instead of ~2 for IMERG).

The use of the coarser WRF24 grid mesh to perform the comparison after conservative remapping can have opposing interpretations. On the one side, the WRF24 model is favored because the fine-scale information produced by the WRF2 and WRF8 models is essentially neglected using the conservative remapping. On the other side, the WRF24 model is somewhat disfavored because its actual effective resolution

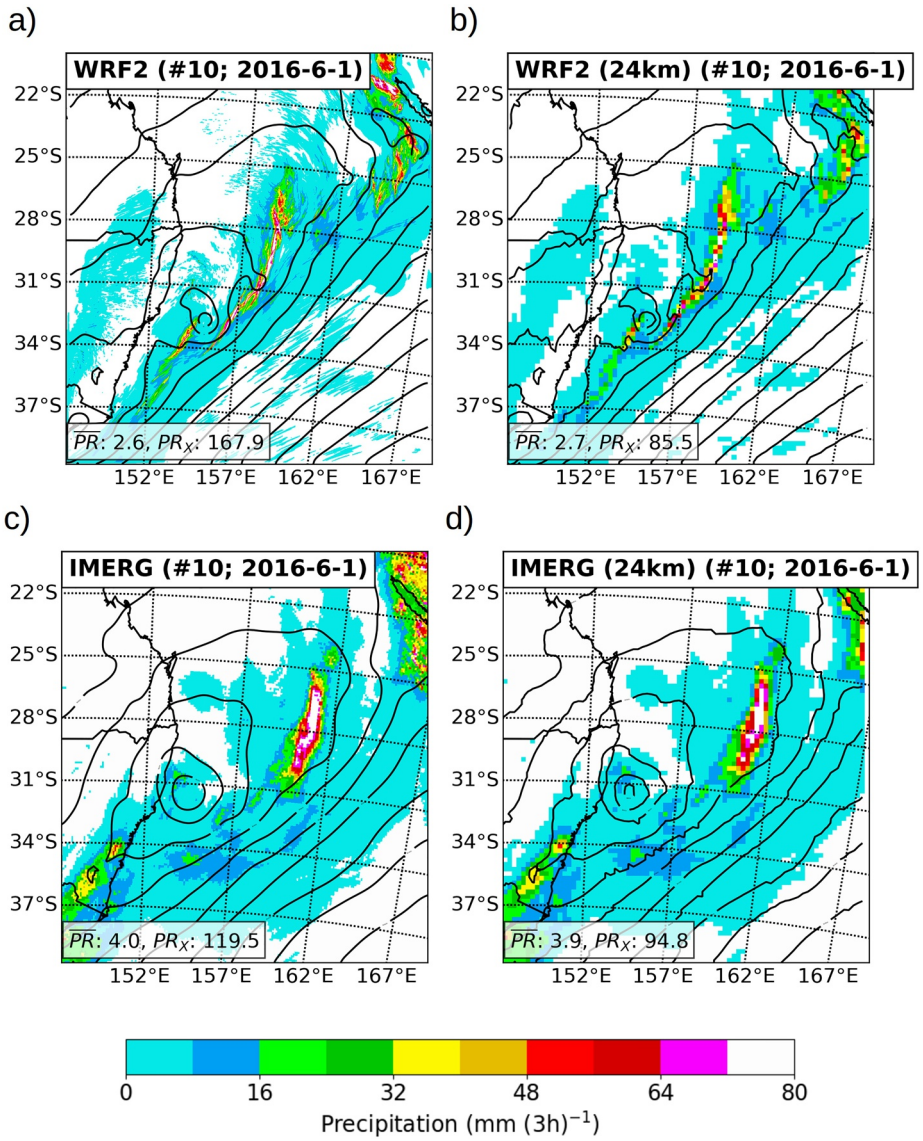

Figure 2. Precipitation rate as simulated in the Weather Research and Forecasting2- control simulation (a and b) and as observed in the Integrated Multi-satellitE Retrievals for GPM product (c and d) for data sets at their native resolution (left panels) and for data sets remapped into the common 24-km grid mesh (right panels). All panels show the precipitation accumulated between 15:00 and 18:00 on the 4th of June 2016. Contour lines show mean sea level pressure from the model in (a and b) and from the ERA5 reanalysis in (c and d). The values $\overline{PR}$ and PR_x in the bottom-left corner of each plot show the domain-average and domain-maximum precipitation rates (in mm (3 h)$^{-1}$), respectively.

is much lower than 24 km leading to numerical errors that are absent when using WR2 and WRF8 (W. C. Skamarock, 2004).

3.2. Environmentally Conditioned Intensity-Frequency (ECIF) Decomposition

In this section, we present a novel approach to decompose precipitation errors that combines a decomposition based on dynamical and thermodynamical environmental regimes and one based on the frequency and intensity of subdaily rainfall events.

Let us first consider a temporal-varying 3-hourly precipitation field, denoted as $P_{x,t}$, that has been regridded into the 24-km common grid mesh as described in Section 3.1. The index x varies between 0 and the total

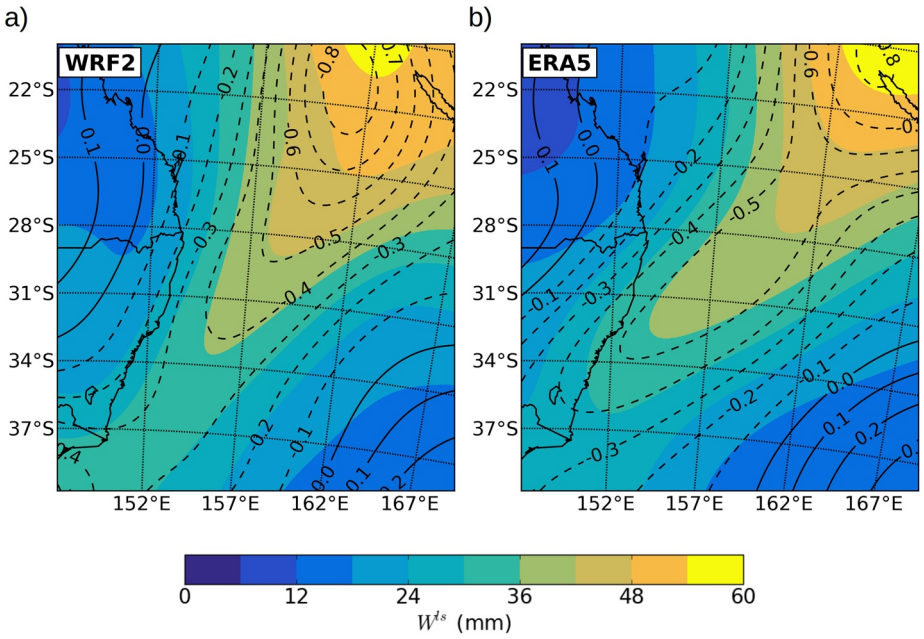

Figure 3. Large-scale 500-hPa pressure vertical velocity ω^{ls} (contour lines; in Pa s^{-1}) and integrated water vapor W^{ls} (colors; in mm) for the Weather Research and Forecasting2-control simulation (a) and the ERA5 reanalysis (b). Both panels show the mean value between 12:00 and 18:00 on the 4th of June 2016.

number of grid points in the region, and t varies between 0 and the total number of 3-hourly data (616) which is given by the number of storm events (11) times the number of time steps per event (i.e., 56 as we discard the first day of the 8-day simulation).

Each 3-hourly precipitation value $P_{x,t}$ has an associated pressure vertical velocity at 500 hPa ($\omega_{x,t}$) and a vertically integrated water vapor ($W_{x,t}$). For observations, ω and W variables are derived from the ERA5 (or ERAI) reanalysis while for simulations both variables are obtained from the simulations themselves. To ensure that vertical motions mostly reflect large-scale circulations and that vertical motions associated with small-scale convection are filtered out, we performed a spatiotemporal smoothing of ω and W fields in both simulations and reanalysis. The smoothing is done by averaging in space ($\Delta x \sim 300$ km) and time ($\Delta t = 6$ h) each local value of $\omega_{x,t}$ and $W_{x,t}$ and we denote the new large-scale variables by $\omega^{ls}_{x,t}$ and $W^{ls}_{x,t}$, respectively. The spatial smoothing is performed using a Gaussian filter with a sigma equal to 300 km.

Figure 3 shows ω^{ls} and W^{ls} for the CTL WRF2 simulation (a) and the ERA5 reanalysis (b) between 12:00 and 18:00 on the 4th of June 2016. In both the simulation and the ERA5 reanalysis, the integrated water vapor W^{ls} varies between 10 and 60 mm and the vertical velocity between −0.8 and 0.2 Pa s^{-1} (0.1 Pa s^{-1} corresponds to about 2 cm s^{-1}). Despite CTL WRF2 having a much higher resolution than ERA5 (about a factor 10), the ω^{ls} and W^{ls} fields appear very similar due to the spatiotemporal smoothing applied. Also, note that the fields ω^{ls} and W^{ls} are strongly correlated with each other and regions with strong ascending motion ($\omega^{ls} \ll 0$) are associated with very humid areas ($W^{ls} \gg 0$).

The dependence of the precipitation on environmental conditions is then assessed by the discretizing (i.e., binning) ω^{ls} and W^{ls} into different categories (that are here denoted as "regimes" as in Bony et al. (2004)) with ω^{ls} between $[\omega^{ls}_i, \omega^{ls}_{i+1}]$ and W^{ls} between $[W^{ls}_j, W^{ls}_{j+1}]$. Based on the values of $\omega^{ls}_{x,t}$ and $W^{ls}_{x,t}$, 3-hourly precipitation rates ($P_{x,t}$) are remapped into the environmental regimes phase space so that $P_{x,t} \rightarrow P_{i,j,n_{i,j}}$. Each regime (i, j) thus contains a total number of events $N_{i,j}$ and a total precipitation ($PT_{i,j}$) given by,

$$PT_{i,j} = \sum_{n_{i,j}=0}^{N_{i,j}} P_{i,j,n_{i,j}} = N_{i,j} \cdot \overline{I}_{i,j} \tag{2}$$

where $\overline{I}_{i,j}$ is the mean precipitation intensity across all events in a given regime. The total precipitation in each regime $PT_{i,j}$ can be approximated as the product of the number of precipitating events ($N_{i,j}^{rain}$) and their average precipitation intensity ($\overline{I}_{i,j}^{rain}$):

$$PT_{i,j} \approx N_{i,j}^{rain} \cdot \overline{I}_{i,j}^{rain} \tag{3}$$

where a 3-hourly event is considered a raining event only if its total precipitation is larger than or equal to 0.25 mm. The choice of this threshold, similar to the one used by Dai et al. (2017) (0.1 mm h^{-1}), leads to an error in the approximation of Equation 3 below 5% of the total precipitation. In addition, the number of precipitating events ($N_{i,j}^{rain}$) can also be expressed as the product of the total number of events ($N_{i,j}$) and the probability of precipitation occurring given the specific environmental conditions ($p_{i,j} = p(rain \mid (\omega_i^{ls}, W_j^{ls}))$):

$$PT_{i,j} \approx N_{i,j} \cdot p_{i,j} \cdot \overline{I}_{i,j}^{rain} \tag{4}$$

where $p_{i,j}$ varies between 0 and 1. Defined this way, $PT_{i,j}$ gives the total amount of precipitation that fell within each regime (i, j) and as such, it depends on the size of the region (i.e., # of grid points) and the length of the period being considered. To compare across regions of different size (e.g., land vs. ocean grid points), we normalize $PT_{i,j}$ by the total number of grid points to define $\widetilde{PT}_{i,j}$ which provides the total precipitation per grid cell.

Using Equation 4, the simulated total precipitation error in each regime is given by the difference between modeled (m) and observed (o) estimations ($\widetilde{PT}_{i,j}^{\epsilon} = \widetilde{PT}_{i,j}^{m} - \widetilde{PT}_{i,j}^{o}$). Replacing using Equation 4 and rearranging terms we have (for simplification we have removed indices (i, j)):

$$\begin{aligned} \widetilde{PT}^{\epsilon} &= \widetilde{PT}^{m} - \widetilde{PT}^{o} \\ &= N^{m} \cdot p^{m} \cdot \overline{I}^{rain,m} - N^{o} \cdot p^{o} \cdot \overline{I}^{rain,o} \\ &= \Delta N \cdot p^{o} \cdot \overline{I}^{rain,o} + \Delta p \cdot \overline{I}^{rain,o} \cdot N^{o} + \Delta \overline{I}^{rain} \cdot N^{o} \cdot p^{o} + R^{\epsilon} \\ &= N^{\epsilon} + p^{\epsilon} + I^{\epsilon} + R^{\epsilon} \end{aligned} \tag{5}$$

where Δ denotes the difference between modeled and observed values (e.g., $\Delta N = N^{m} - N^{o}$). Equation 5 explicitly states that errors in the total precipitation in each regime (i, j) can result from four distinct terms:

1. N^{ϵ} is associated with the ability of the model to simulate the observed frequency of specific large-scale environmental conditions ("regimes") and is zero only when the model simulates the right frequency of occurrences $N^{m} = N^{o}$. This error is denoted as the "environment" component of the error.
2. p^{ϵ} is associated with the ability of the model to simulate the observed number of precipitation events given some environmental conditions and is zero only when the simulated probability of precipitation occurrence is as observed $p^{m} = p^{o}$. This error is denoted as the "frequency" component of the error.
3. I^{ϵ} is associated with the ability of the model to produce the right rainfall rate given some environmental conditions and is zero when $\overline{I}^{rain,m} = \overline{I}^{rain,o}$. This error is denoted as the "intensity" component of the error.
4. R^{ϵ} is associated with interaction terms and is denoted here as the "residual" component of the error. It is shown to be small compared with the other errors. R^{ϵ} can be expressed as:

$$R^{\epsilon} = \Delta N \cdot \Delta p \cdot \overline{I}^{rain,o} + \Delta p \cdot \Delta \overline{I}^{rain} \cdot N^{o} + \Delta \overline{I}^{rain} \cdot \Delta N \cdot p^{o} + \Delta p \cdot \Delta \overline{I}^{rain} \cdot \Delta N \tag{6}$$

In any given regime, the calculation of the error using Equation 5 is well defined when both (or when neither) simulations nor observations show some precipitating events. In the case in which only one of the two data sets show some precipitating events the error cannot be calculated using Equation 5. In such a case, the error is defined as the value of the precipitation in either data set, with a minus sign when the precipitation is missing in the model and a positive sign when the precipitation is missing in the observations. In these cases, we assume that the error is associated with the large-scale environmental component of the error (N_{ϵ}). This way, the approach might be more insightful if the method is going to be applied to look at future climate simulations (future changes instead of errors) where large differences will arise due to differences

in the environment conditions (e.g., the amount of precipitable water). Note that for regimes that occur without any precipitation event, variables $N_{i,j}$ and $p_{i,j}$ will be well defined but the intensity $\overline{I}^{rain}$ will not.

3.3. Error Metrics Including Absolute and Additive Absolute Errors

It is easy to see from Equation 5 that for any given regime, small total precipitation errors can be obtained by compensating positive and negative errors across different error terms. Moreover, it is also easy to see that a small overall precipitation error could be obtained by compensating positive and negative errors across different regimes. To assess the role played by the compensation of errors and whether models simulate the right precipitation for the right reasons, we define four new error metrics that allow a different degree of error compensation.

The first error metric, denoted as the absolute error (AE), allows for all possible error compensations (i.e., compensations across regimes and across error terms) and is calculated as the absolute value of the sum of all errors:

$$PT_{\epsilon}^{AE} = \mid \sum_{i,j} \widetilde{PT}^{\epsilon}{}_{i,j} \mid = \mid \sum_{i,j} (N_{i,j}^{\epsilon} + p_{i,j}^{\epsilon} + I_{i,j}^{\epsilon} + R_{i,j}^{\epsilon}) \mid \tag{7}$$

The second error metric, denoted as the regimes additive AE (R-AAE), prevents the compensation of errors occurring between different environmental regimes and is calculated as the sum of AEs of the total precipitation in each regime (i, j):

$$PT_{\epsilon}^{R-AAE} = \sum_{i,j} \mid \widetilde{PT}^{\epsilon}{}_{i,j} \mid = \sum_{i,j} \mid N_{i,j}^{\epsilon} + p_{i,j}^{\epsilon} + I_{i,j}^{\epsilon} + R_{i,j}^{\epsilon} \mid \tag{8}$$

The third error metric, denoted as the environment-frequency-intensity-additive AE (EFI-AAE), prevents the compensation of errors occurring between the environment, frequency and intensity error terms and is calculated as follows:

$$PT_{\epsilon}^{EFI-AAE} = \mid \sum_{i,j} N_{i,j}^{\epsilon} \mid + \mid \sum_{i,j} p_{i,j}^{\epsilon} \mid + \mid \sum_{i,j} I_{i,j}^{\epsilon} \mid + \mid \sum_{i,j} R_{i,j}^{\epsilon} \mid \tag{9}$$

Finally, the fourth error metric, denoted as the additive AE (AAE), prevents all possible error compensations within the framework and is calculated as the sum of AEs for individual regimes and error terms:

$$PT_{\epsilon}^{AAE} = \sum_{i,j} \mid N_{i,j}^{\epsilon} \mid + \sum_{i,j} \mid p_{i,j}^{\epsilon} \mid + \sum_{i,j} \mid I_{i,j}^{\epsilon} \mid + \sum_{i,j} \mid R_{i,j}^{\epsilon} \mid \tag{10}$$

Defined this way, the four error metrics provide a hierarchy of errors allowing for different degrees of error compensation. The PT_{ϵ}^{AAE} is always greater or equal to the other errors while the PT_{ϵ}^{AE} is smaller or equal to the other errors. In particular, PT_{ϵ}^{AAE} is zero only when errors in all individual regimes and terms are zero while the other three errors might be zero even if individual errors are non-zero but compensate with each other. Note that from a purely mathematical perspective, there are many simulated rainfall fields that can lead to a zero error in the case of PT_{ϵ}^{AE} but only one that can lead to a zero error in PT_{ϵ}^{AAE}.

Model errors are calculated relative to the median across all five satellite-based products as the reference data. The observational uncertainty can then be estimated by calculating error metrics using individual satellite-based products relative to the median across all products ($PT_{\epsilon,obs_i}^{AAE}$). For example, the observational range can be estimated as the largest error across all five satellite-based products obs_i:

$$PT_{\epsilon,obs-5}^{AAE} = \max_{i}(PT_{\epsilon,obs_i}^{AAE}) \tag{11}$$

A second observational uncertainty measure can be obtained by considering the second largest error to estimate $PT_{\epsilon,obs-4}^{AAE}$. Error metrics are sometimes normalized by the median observed total precipitation across all regimes: $\Sigma_{i,j} \widetilde{PT}^{o}{}_{i,j}$.

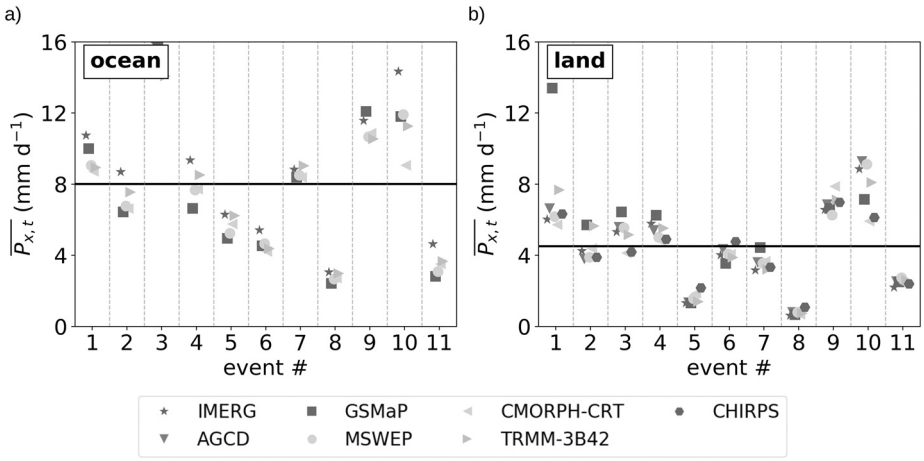

Figure 4. Domain-averaged precipitation rates (in mm d^{-1}) for individual events over land (a) and ocean (b) grid points for several observational products (see colors/symbols). The black horizontal line shows the precipitation averaged across observed-based products, events and the domain. AWAP and CHIRPS are only available over land grid points.

4. Results

4.1. Assessing Total Precipitation Biases

Averaged precipitation from observations (gray markers) over ocean grid points and individual events is shown in Figure 4a. Averaged precipitation shows large differences across events with values varying between about 2 mm d^{-1} (#8, 2008-8-18) and about 12 mm d^{-1} (#10, 2016-6-1). There is quite strong agreement among observational products with mean precipitation differences that are usually within 15%. However, there are some systematic differences with IMERG usually providing the highest mean precipitation rate (about 1 mm d^{-1} larger than the other data sets) and GSMaP the lowest precipitation rate among all five observational products.

Results for land grid points (Figure 4b) are qualitatively similar to those for ocean grid points but averaged rainfall rates are lower by about 40%. The observational range over land grid points is similar to the one obtained over ocean with the exception of the 2007-6-4 (#1) event for which GSMaP estimation is twice as large as any other estimation. Over land, all satellite-based products compare well with the gridded station data set (AGCD) and difference are again usually within 15%.

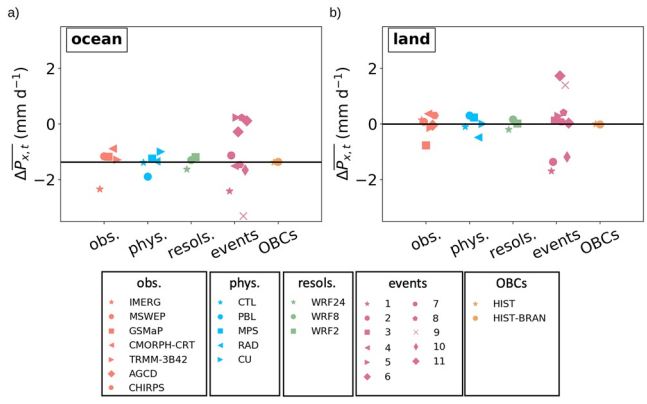

Figure 5. Mean biases for different subensembles over land (a) and ocean (b) grid points. Different subensembles are obtained by considering the different dimensions of the full ensemble (resolutions (3), physics (5), events (11), and oceanic boundary conditions (2); see Equation 1) and by fixing, one at the time, the different values of each dimension. In addition, the column "obs." shows the bias calculated separately for each observational product. The black horizontal line shows the averaged precipitation bias across the full ensemble.

Figure 5 shows precipitation mean biases for several subensembles of simulations relative to the mean across all observations. Values in Figures 5a and 5b show mean biases calculated by averaging over the full ensemble while leaving constant the specific factor considered. For example, the resolutions column ("resols.") shows three values, one for each resolution, that were obtained by averaging precipitation rates across simulations with a given resolution. Similarly, the observations column ("obs.") shows five and seven values in panels (a) and (b), respectively, showing mean biases calculated using individual observations instead of the mean across all observations (i.e., the "obs" column gives an idea of the observational uncertainty). Results show that the full ensemble has a negative bias of about −1.2 mm d^{-1} over the ocean and a nearly zero bias over land. WRF24 simulations seem to produce about 0.5 mm d^{-1} less precipitation than WRF2 simulations regardless of whether the average is calculated over ocean or land grid points (see "resols". column). The choice of the model physics ("phys") also leads to some differences in mean biases and the magnitude and sign of these biases depend on

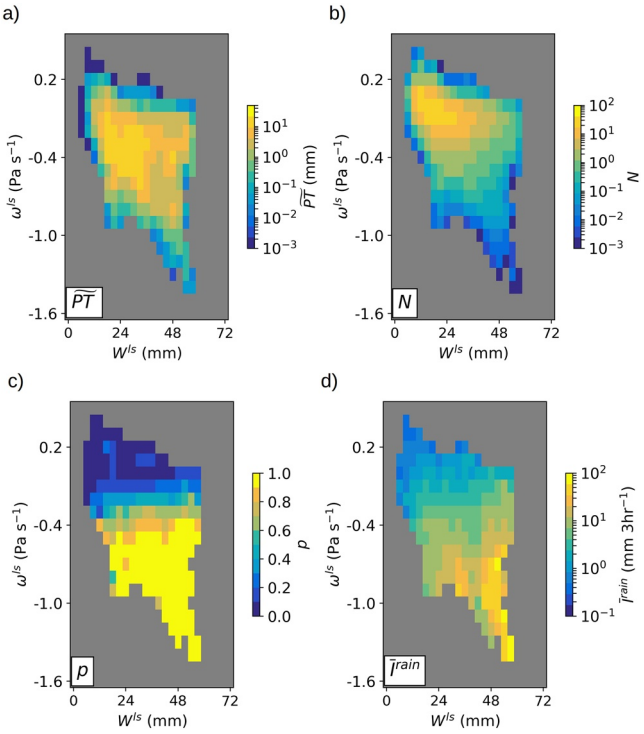

Figure 6. Panel (a) shows the Integrated Multi-satellitE Retrievals for GPM total precipitation per grid point for different regimes of integrated water vapor (W^{ls}; every 3 mm) and 500-hPa pressure vertical velocity (ω^{ls}; every 0.1 Pa s^{-1}) as obtained from the ERA5 reanalysis. The total precipitation is decomposed according to the number of times that environmental conditions occurred in each regime (b), the probability of precipitation occurring in each regime (c) and the intensity of precipitation in each regime (d). Terms are calculated using 3-hourly data and for regimes with at least five precipitating events by pooling together precipitation from all events. See Section 3.2 for details on the decomposition.

whether ocean or land grid points are considered. For example, changing the surface and planetary boundary layer scheme from YSU (CTL member) to MYJ (PBL member) leads to more precipitation over land but less precipitation over the ocean. Also, the change of the cumulus scheme from BMJ (CTL) to KF (CU) has a larger effect over the ocean than over land. The use of different OBCs (HIST vs. HIST-BRAN) seems to have very little effect on the total precipitation with both subensembles leading to nearly the same mean biases. Finally, column "obs." shows that the choice of the observational product use to calculate the bias can have a substantial impact on the value of the ensemble mean bias. Over ocean grid points, the largest negative biases are found compare with IMERG data set while over land grid points biases are largest compared with the GSMaP product.

4.2. Decomposing Precipitation Over Ocean

Figure 6a shows IMERG total precipitation per ocean grid point for regimes obtained using large-scale 500-hPa omega (ω^{ls}) and vertically integrated water vapor (W^{ls}) from the ERA5 reanalysis. As shown in Equation 4, in each regime the total precipitation can be described as the product of three terms: the number of occurrences in each regime (Figure 6b), the probability of precipitation occurring in each regime (Figure 6c) and the mean intensity of precipitating events in each regime (Figure 6d). Figures S3 and S4 show the total precipitation in each regime over ocean grid points for the five different observation products (Figure S3) and for the CTL-HIST member for the three resolutions WRF24, WRF8, and WRF2 (Figure S4).

Total precipitation (Figure 6a) is distributed quite uniformly across a wide range of regimes with values peaking at about 20 mm for regimes with $\omega^{ls} \sim -0.4$ Pa s^{-1} and $W^{ls} \sim 28$ mm. However, decomposition terms (Figures 6b–6d) show that the total amount of precipitation is attained differently for different regimes. The most frequent regimes (N; Figure 6b) are those with $\omega^{ls} \sim 0$ Pa s^{-1} and $W^{ls} \sim 18$ mm and the number of occurrences decreases rapidly as we consider regimes with stronger dynamical forcing and/or more humid environments. On the contrary, the probability of precipitation being triggered (p (rain|N); Figure 6c) and the mean intensity ($\overline{I}^{rain}$; Figure 6d) both increase as we consider stronger dynamical forcing and more humid environments. According to IMERG observations, the probability of precipitation being triggered is nearly 0 for regimes with $\omega^{ls} \geq 0$ Pa s^{-1} and nearly 1 for regimes with $\omega^{ls} < -0.5$ hPa s^{-1}. The mean intensity of rainfall attains about 100 mm every 3 h for very humid environments. In addition, Figures 6a–6d show that strongest ascending motions only occur simultaneously with very large values of W^{ls} leading to an elongated shape in the occurrence of regimes that suggests a strong relation between ascending motion and low-level moisture advection processes around cyclonic circulations.

Figure 7 shows the total precipitation bias averaged across WRF2 simulations for individual regimes (Figure 7a) and the three error decomposition terms N_ϵ, T_ϵ, and I_ϵ (Figures 7b–7d). The contribution to the total error by the residual term is small and is shown in Figure S5. Biases are calculated relative to the median across the five observational products. The stippling (see "x" markers) in the total error and the frequency and intensity error terms (Figures 7a and 7c–7d, respectively) denotes those regimes where the sign of the bias agrees across the five observational products.

WRF2 simulations overestimate the total precipitation for regimes with low W^{ls} and weak ω^{ls} and underestimates the precipitation for most other regimes (Figure 7a) with all three decomposition terms (Figures 7b–7d) contributing substantially to the total precipitation error. According to the environment er-

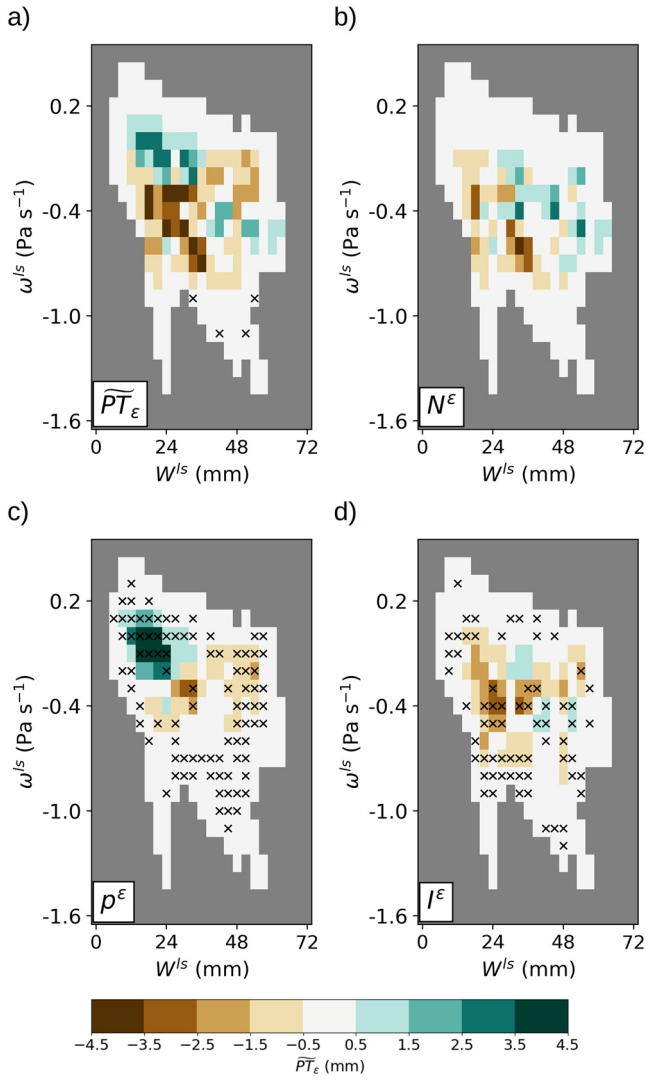

Figure 7. Weather Research and Forecasting2-control error relative to the median across five satellite-based observations (CMORPH-CRT, GSMaP, IMERG, MSWEP and TRMM-3B42) for the total precipitation per grid point (a) and the three decomposition terms: environment error (b), frequency error (c), and intensity error (d). Results are for ocean grid points. In panels (a, c, and d), "x" markers denote those environmental regimes where the sign of the bias agrees among the five observations. All terms are calculated using 3-hourly data and for regimes with at least five precipitating events. See Section 3.2 and Equation 5 for details on the error decomposition.

ror term (Figure 7b), WRF2 produces a negative bias for low W^{ls} regimes and a positive bias for high W^{ls} regimes suggesting that WRF2 produces moister environments than the ERA5 reanalysis over most dynamical regimes. The frequency error term (Figure 7c) shows that WRF2 simulations produce precipitation too often for regimes with low W^{ls} and weak ω while they produce precipitation too rarely for regimes with high W^{ls} and moderate ω. As shown by the stippling, these biases are robust to the choice of the observational data set. The intensity error term (Figure 7d) shows that WRF2 simulations underestimate the intensity of precipitation over most regimes. Again, this result is often robust to the choice of observations.

Total precipitation errors as a function of W^{ls} averaged across weak ($\omega \geq -0.3$ Pa s^{-1}) and strong ($\omega^{ls} < -0.3$ Pa s^{-1}) dynamical regimes are shown in Figure 8 for individual simulations. For weak dynamical regimes, the total precipitation error (Figure 8a) is usually larger than zero for low values of W^{ls} ($W^{ls} \leq 30$ mm) and smaller than zero for high values of W^{ls} ($W^{ls} \geq 40$ mm) although only biases for high values of W^{ls} seems to be robust to the choice of the observed data set. Looking at individual decomposition terms, it is seen that the small errors for low-to-moderate W^{ls} values result from a compensation between robust positive errors in the frequency term (Figure 8e) and negative errors in the intensity term (Figure 8g). This result is valid for most individual members of the ensemble although WRF2 simulations show smaller errors for both the frequency and intensity terms. For high values of W^{ls} and weak dynamical regimes, WRF2 members show a distinct behavior compared with WRF8/WRF24 with opposite error signs for the frequency term and smaller errors for the intensity term. Errors in the environment error term show a modest underestimation of low W^{ls} and overestimation for moderate and high W^{ls} consistent with all simulations showing a positive bias in the W^{ls}.

Results for strong dynamical regimes (right panels in Figure 8) show that most members underestimate the total precipitation for all W^{ls} values, largely due to a negative error in the environment term and to a lesser extent, in the intensity term. Errors in the environment term are much larger than for weak dynamical regimes suggesting that the magnitude of the W^{ls} bias depends on the environmental conditions. Figure 8b also shows that the bias in W^{ls} depends on the choice of resolution and physics with errors that are smallest among WRF2 simulations and largest among WRF24 simulations. Errors in the intensity and frequency terms are smaller for strong than for weak dynamical environments and differences between resolutions are also usually smaller.

Overall, the choice of the physics seems to have a modest impact in most errors although there are some exceptions. For all dynamical regimes, the environment error term is mostly affected by the representation of boundary layer processes (CTL vs. PBL) and to a lesser extent, by the representation of convective processes (CTL vs. CU), particularly for WRF24 and WRF8 for strong dynamical regimes. For weak dynamical regimes, the frequency error term is mostly affected by the representation of convective processes, even when considering WRF2 simulations presumably because of the effect through the boundary conditions. The intensity error term is again mostly affected by the choice of the cumulus scheme in the case of WRF8 and WRF24. For WRF2 simulations, the choice of the MPS and PBL schemes seems to play the largest role for weak and strong dynamical regimes, respectively.

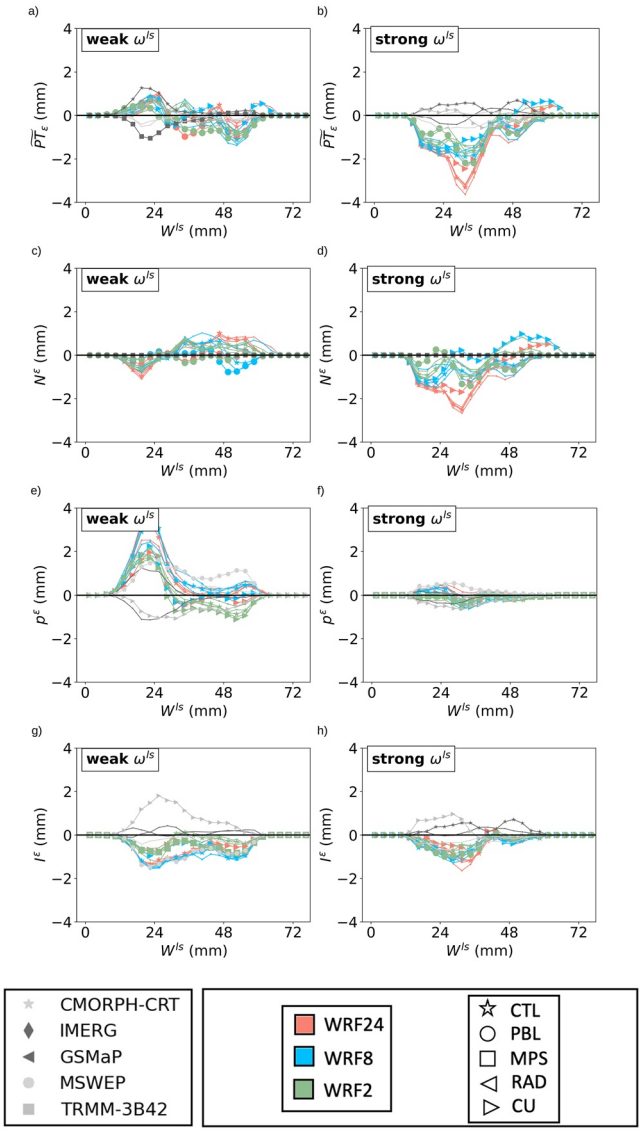

Figure 8. Simulated errors relative to the median across observations as a function of integrated water vapor (W^{ls}) for weak ($\omega^{ls} \geq -0.3$ Pa s^{-1}; left panels) and strong ($\omega^{ls} < -0.3$ Pa s^{-1}; right panels) dynamical regimes. Results are shown for errors in the total precipitation (a and b), the environment (c and d), the frequency (e and f), and the intensity (g and h) terms. Colors show simulations performed using different resolutions and markers show different physics members. For each resolution and physics, the average across the two oceanic boundary conditions members is shown. Gray colors show errors for individual observations. The marker size is largest for the control (CTL) and the physics scheme leading to the largest difference compare to the CTL run. All terms are calculated using 3-hourly data for ocean grid points and for regimes with a minimum of five precipitating events across all storms.

4.3. Decomposing Precipitation Over Land

IMERG total precipitation and decomposition terms per land grid point are shown in Figure 9. The range of regimes over land grid points decreases compared with ocean grid points due to a lesser occurrence of regimes with high or very high W^{ls} and ω^{ls}. In addition, the intensity of precipitation is usually higher over ocean compared with land grid points when considering regimes with the same environmental conditions (i.e., same ω^{ls} and W^{ls}). As a consequence, the average total precipitation over land grid points is lower than over ocean grid points (as shown in Figure 4).

The near-zero error shown in the domain-averaged precipitation plot (Figure 4b) results from the compensation between positive and negative errors across different regimes and across error terms (Figure 10). Similar to ocean grid points, the environment error shows that WRF2 simulations are usually too humid compared with observations. WRF2 simulations systematically overestimate the frequency of precipitation (Figure 10c) and underestimate the mean intensity of precipitating events (Figure 10d) compared with observations regardless of the choice of the reference data set.

For weak dynamical regimes (left panels in Figure 11), errors in total precipitation are positive and generally small although the spread across individual members is quite large. Errors in the environment term are similar in WRF2 and WRF24 simulations while WRF8 simulations show very large errors when using the default cumulus scheme (BMJ). This suggests that some of the hypotheses used in the BMJ cumulus scheme, that seem to work relatively well for the WRF24 simulation, fail in the "gray zone" for the WRF8 simulation. Moreover, it seems that some of the large errors in WRF8 simulations using BMJ cumulus scheme are inherited by WRF2 simulations through the boundary conditions and WRF2 simulations also show the lowest errors for the CU simulation that uses the KF scheme instead of the BMJ scheme. All simulations show a clear compensation between large positive errors in the frequency term and negative errors in the intensity term.

For strong dynamical regimes (right panels in Figure 11), errors in total precipitation are mostly negative and largely explained by errors in the environment term. As for ocean grid points, WRF2 simulations outperform WRF8 and WRF24 particularly by improving the representation of the environment term. Errors in the frequency and intensity terms are relatively small and do not show the strong compensation found for weak dynamical regimes.

The effect of changing the representation of subgrid-scale processes differs for the various error terms. For the environment error, the largest differences compared to the CTL run are usually related with the choice of the PBL and CU schemes, depending on the resolution of the simulation. For all resolutions, the frequency error term is mostly affected by the choice of the PBL scheme for weak dynamical regimes and the choice of the MPS scheme for strong dynamical regimes. Finally, the intensity error term is generally little affected by choices in different physics although the choice of the CU scheme leads to substantial differences for WRF24 and WRF8 simulations for strong dynamical regimes.

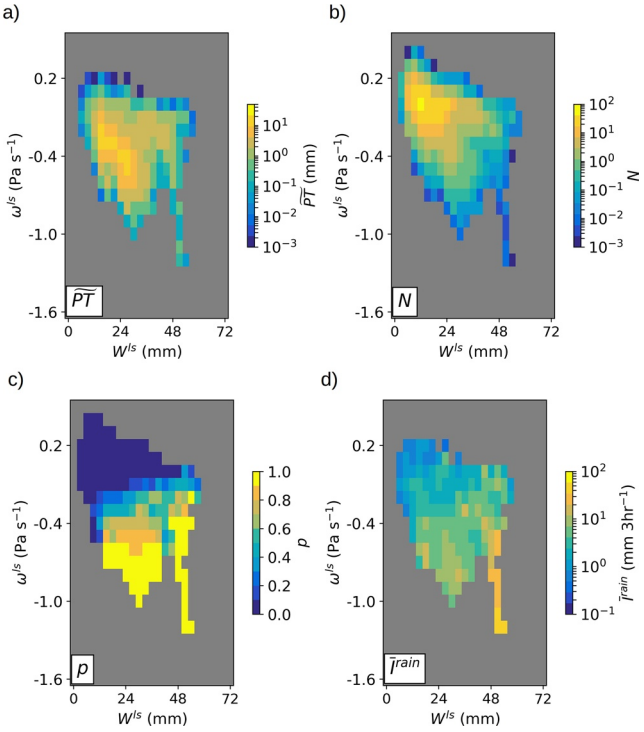

Figure 9. As in Figure 6 but for land grid points.

Figures S6 and S7 show that results are qualitatively very similar when calculating the observed regimes using the ERAI instead of the ERA5 reanalysis. The bias in the W^{ls} is also present when comparing with the ERAI reanalysis but appears to be a little smaller.

4.4. Quantifying the Compensation of Errors

Previous sections show that simulations often produce small total precipitation errors by compensating positive and negative errors across different regimes or across different terms. In this section, we quantify the extent to which errors compensate in the different simulations by calculating the four error metrics defined in Section 3.3.

Top panels in Figure 12 show AAEs (PT_{ϵ}^{AAE}; see Equation 10) as a function of AEs (PT_{ϵ}^{AE}; see Equation 7) for individual simulations and observations for ocean (Figure 12a) and land (Figure 12b) grid points. All errors have been calculated relative to the median across the five satellite-based products. There are some fundamental differences between both errors:

1. Error magnitude: additive absolute errors (AAEs) are about six times larger than absolute errors (AEs).
2. Error sensitivity: AAE and AE show a very different sensitivity to the choice of the member. For example, over the ocean, the simulation that shows the best performance according to the AE (WRF8-CU-HIST) only shows an average performance according to the additive error. The level of agreement between the sensitivity of AAE and AE is estimated using the Spearman correlation coefficient (see r_s value in top-left corners of each plot) that shows values of 0.48 and 0.05 over ocean and land grid points, respectively. This suggests that, particularly over land grid points, both errors provide a very different view of the skill of the ensemble.
3. Error information: AAEs are much more efficient than AEs at separating observations from simulations. Over both land and ocean grid points, the AAE of every single model simulation lies outside the observational uncertainty meaning that AAEs are always robust to the choice of the reference data set. On the other hand, the AE of individual simulations is very often within the observational uncertainty over both land and ocean grid points.

Middle panels in Figure 12 show AAEs as a function of the EFI-AAE ($PT_{\epsilon}^{EFI-AAE}$; see Equation 9). Again, PT_{ϵ}^{AAE} is larger than $PT_{\epsilon}^{EFI-AAE}$ and the sensitivity of both errors is similar over ocean ($r_s = 0.88$) but quite different over land grid points ($r_s = 0.52$). In contrast, errors obtained using the environmental regime's AAEs (PT_{ϵ}^{R-AAE}; see Equation 8) are nearly perfectly correlated with the AAEs over land ($r_s = 0.94$; Figure 12f) and well correlated over ocean grid points ($r_s = 0.78$; Figure 12e).

Total AAEs (full bars) and the contribution from individual error terms (in colors) averaged across different subensembles for ocean (left panels) and land (right panels) grid points are shown in Figure 13. In relative terms, AAEs are larger over land than over ocean grid points. The comparison across additive errors in different resolution subensembles (Figures 13a and 13b) shows that WRF2 simulations outperform WRF8/WRF24 simulations by about 20% over both ocean and land grid points. Improvements arise from a reduction of all three error terms dominated by the environment and intensity error terms over ocean grid points and by the environment and frequency error terms over land grid points. Over ocean grid points, simulated AAEs are always larger than the observational uncertainty range (full red line; see Equation 11) suggesting that AAEs are robust to the choice of observational data set. Over land grid points, the full observational uncertainty range is nearly as large as the additive errors but the large observational range arise mostly from the inclusion of the TRMM-3B42 data set and when removed the observational range is much lower (dashed red line).

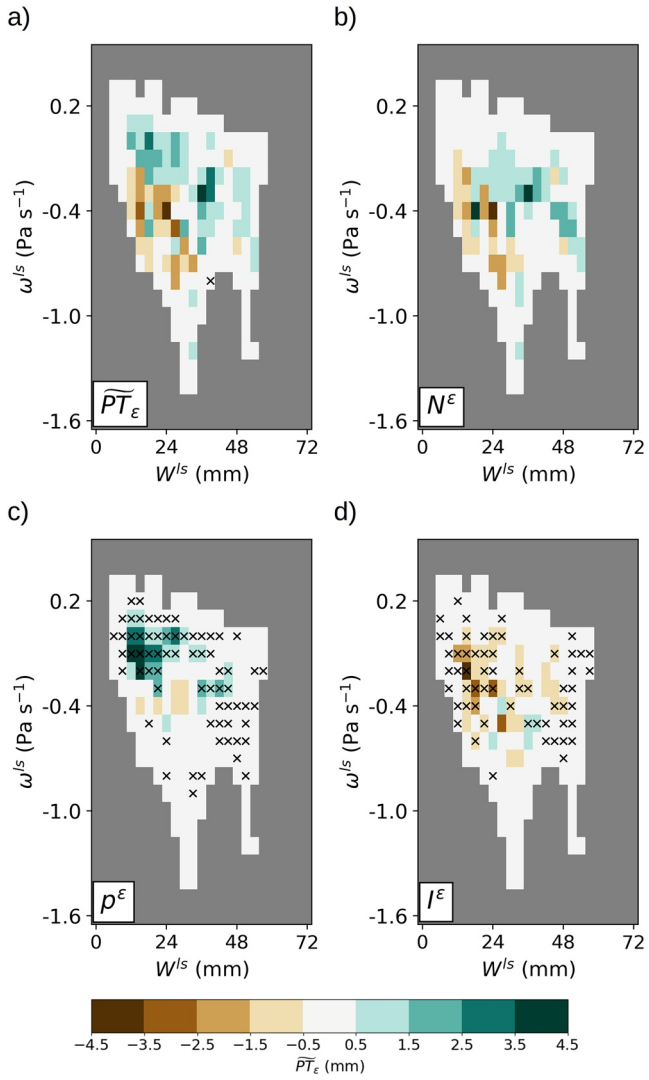

Figure 10. As in Figure 7 but for land grid points.

Differences between total AAEs in different physics subensembles are within 10% over ocean and land grid points (Figures 13c and 13d). Over ocean grid points, the PBL and CU subensembles perform best and improve the CTL subensemble by nearly 10% on average. However, the specific way in which individual physics members improve upon the CTL run differs, with the PBL improving the environment and frequency terms and the CU improving the frequency and intensity terms. Over land grid points, the PBL, RAD, and CU subensembles improve CTL simulations by about 5%. In this case, all three physics members' improvements arise from a better representation of the environment term compared to the CTL run. Note that, while overall improvements for different physics members are modest (i.e., within 10%), these arise from positive and negative error differences in individual error terms.

Finally, differences between subensembles performed with different ocean boundary conditions (Figures 13e and 13f) are negligible over both land and ocean grid points although the subensemble using the highest resolution SSTs (HIST-BRAN) shows a slight improvement over the ocean.

Results obtained using the ERAI instead of the ERA5 reanalysis are shown in Figure S9. When using ERAI, errors are overall smaller and the improvements by the WRF2 subensemble are about 10% instead of 20%.

5. Discussion

We have assessed how well an ensemble of high-resolution simulations represent the precipitation associated with some of the strongest cyclones that affected the east coast of Australia in the last 20 years. The ensemble comprises a total of 330 simulations including simulations performed using different horizontal resolutions, different physics and different ocean boundary conditions. To assess the performance, we decompose observed and simulated 3-hourly precipitation fields using two different but complementary decompositions. The first one involves separating rainfall rates according to dynamical (vertical velocity at 500 hPa) and thermodynamical (water vapor content) regimes. The second involves separating rainfall values in each regime into contributions from the frequency and intensity of precipitation. Using both decompositions, we then describe the error in total precipitation as the sum of three error terms that include errors in large-scale environmental conditions, in the frequency of precipitation and in its intensity.

Our assessment shows a number of systematic errors that are common to all members of the ensemble, regardless of their horizontal resolution, physics schemes, and ocean boundary conditions used, and whether convection is parameterized or explicitly represented:

1. Simulated errors are larger for weak than for strong dynamical regimes particularly when considering the frequency and intensity error terms.
2. Simulations produce precipitation too often with an intensity that is too weak, particularly under weak dynamical forcing. This links with the well-known issue of climate models producing too much drizzle (e.g., Sun et al., 2006) and shows that the higher resolution and the explicit representation of convection do not fully solve this issue.
3. Simulations usually have a substantial bias in the vertically integrated water vapor although the magnitude and sign of this bias depend on the strength of the dynamical forcing.
4. Relative to the total amount of precipitation, errors are larger over land than over ocean grid points mainly due to larger errors in the representation of the environment error term.

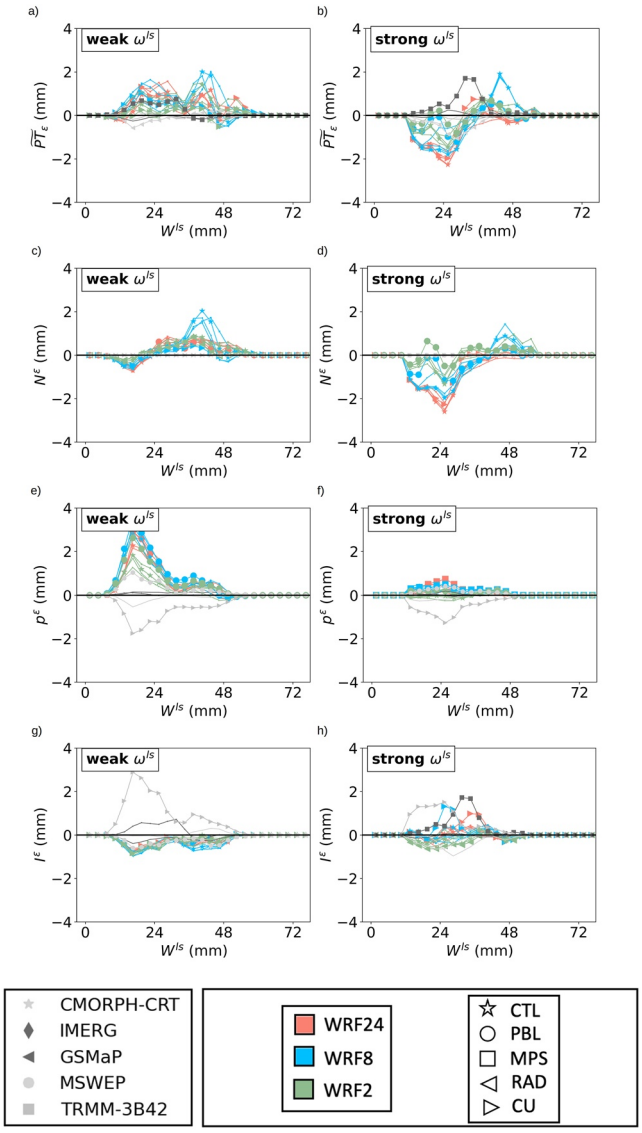

Figure 11. As in Figure 8 but for land grid points.

Our results also show that several aspects of the precipitation field depend strongly on the configuration of the model. Two-km resolution simulations (WRF2), performed without parameterized convection, show improvements in all three error terms compared to WRF8/WRF24. For weak dynamical regimes, improvements are dominated by smaller errors in the frequency and intensity terms while for strong dynamical regimes, errors are dominated by the environment error term. Specifically, improvements in the environmental conditions are related to a smaller positive bias in the vertically integrated water vapor likely due to a better representation of evaporation processes and the vertical transport of moisture. Further work is, however, needed to better understand these links and the moisture differences between WRF2 and WRF8/24 including, for example, the influence of increasing the vertical resolution. In addition, mostly for weak dynamical regimes, WRF2 improves the simulation of the frequency and intensity of rainfall compared with WRF8/WRF24. Over land, WRF8 simulations perform poorer than WRF24 due to large errors in the representation of large-scale environmental conditions that are very sensitive to the choice of the cumulus scheme. We speculate that this is due to WRF8 operating in the so-called "gray zone".

The choice of subgrid-scale physics schemes has a strong impact on the total error mostly by affecting the error in large-scale environmental conditions. In this regard, over both land and ocean grid points, the main differences arise when changing the cumulus and planetary boundary layer schemes and these differences are mostly explained by the total amount of moisture made available, either through the vertical transport or through evaporation processes. These results are in agreement with Gilmore et al. (2016). Overall simulations are little affected by changes in the spatial resolution of the ocean boundary conditions (i.e., SSTs).

A key aspect of the assessment is the use of several high-spatiotemporal resolution rainfall products in order to evaluate the observational uncertainty. Our results indicate that differences between observational products are large but when combining information from individual decomposition terms, observational products agree quite well with each other and provide a fundamentally different picture compared to simulations. In this regard, among the five satellite products used in this study, the largest contribution to the observational range over ocean grid points arise from the TRMM-3B42 and MSWEP products. Over land grid points, the inclusion of TRMM-3B42 produces a doubling of the observational range suggesting some fundamental difference with IMERG, MSWEP, CMORPH-CRT, and GSMaP. In addition, the calculation of the observed large-scale environmental regimes requires the use of reanalysis data. To assess the dependence on the choice of reanalysis, we have calculated the observed regimes based the ERA5 and the ERAI reanalyzes and our results show little dependence to this choice.

Using the three error terms, we define several new error metrics that quantify the amount of compensation between errors of different sign and provides a completely different view of the skill of the ensemble compared with the commonly used bias or AEs. The new AAE provides a better characterization of the overall performance because:

1. AAE penalizes models that produce the right results for the wrong reasons making this metric more resistant to model calibration and tuning.
2. AAE is much more efficient than the AE at separating observations from simulations. This result shows the value of breaking-up errors into components as it is a much more difficult task for models to correctly get all individual components than to correctly get the final value.

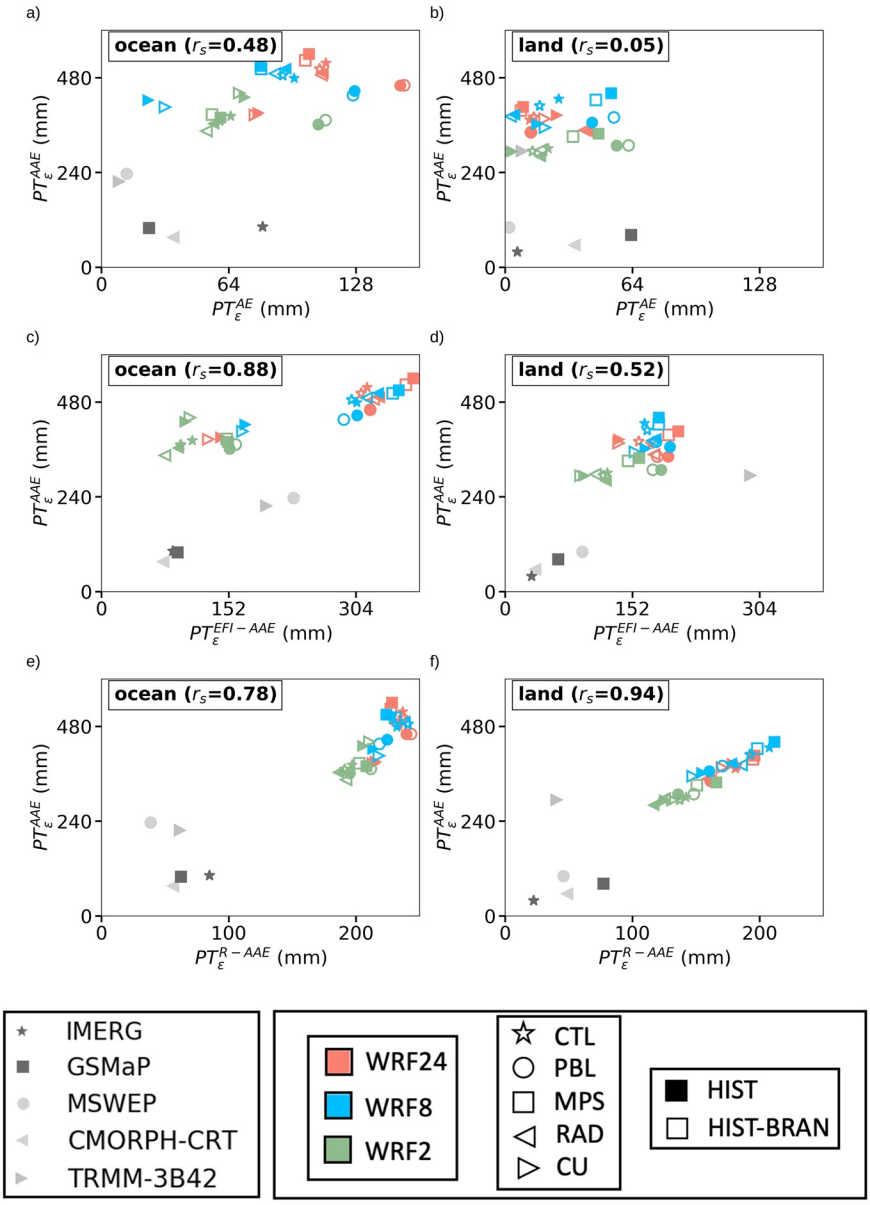

Figure 12. Additive absolute errors as a function of absolute (top panels), intensity-frequency-additive errors (middle panels), and regimes additive errors (bottom panels). Left panels show results for ocean grid points and right panels for land grid points. All errors are calculated relative to the median across the five observations. Gray colors show errors for individual observations and other colors for individual simulations. The r_s value in the top-left corner shows the Spearman correlation coefficient between simulated errors. Refer to Section 3.3 for the definition of errors.

3. Contrary to the AE, AAE is never close to zero and differences across errors in individual members of the ensemble are much smaller than the errors themselves. We speculate that this is a more realistic behavior of the actual errors.

6. Conclusions

We have presented a decomposition of total precipitation, denoted as Environmentally Conditioned Intensity-Frequency (ECIF), that can be equally applied to simulations and observations and can be used to evaluate the ability of climate models to represent precipitation processes in an insightful way. We argue that this

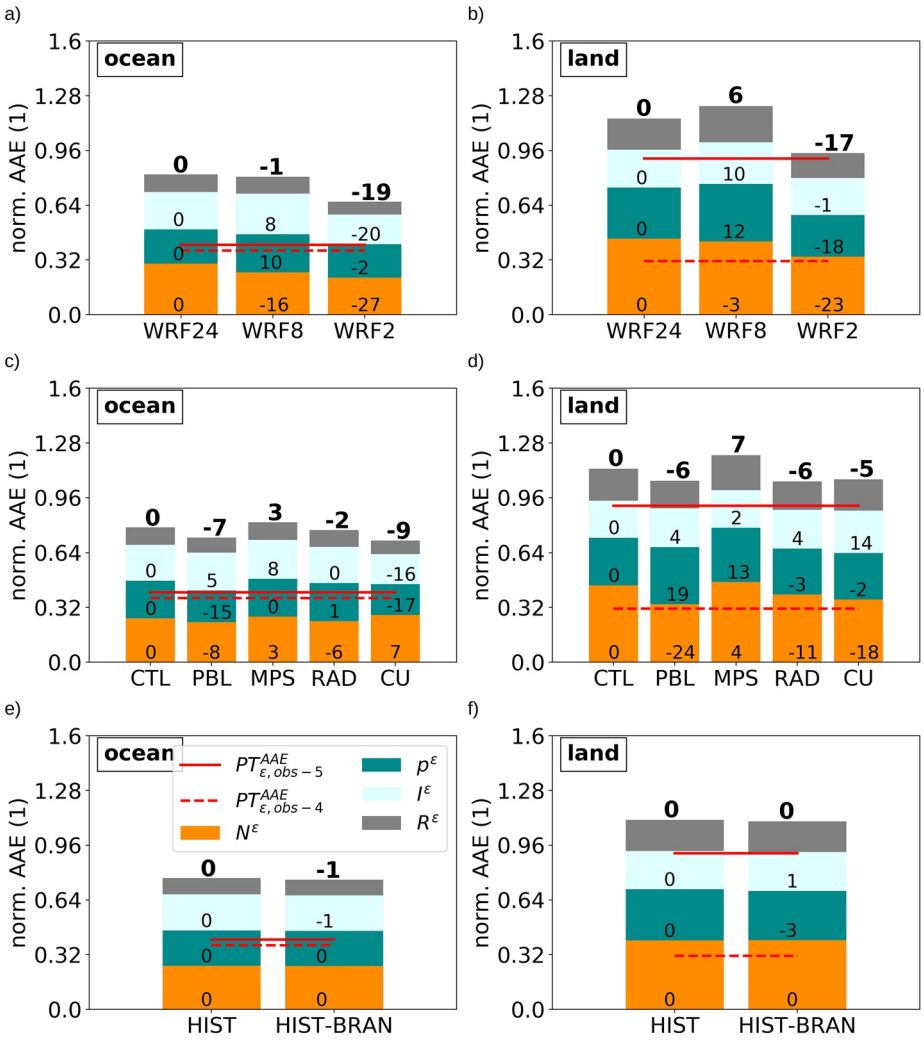

Figure 13. Normalized precipitation additive absolute errors (full bars) and individual error terms (colors bars) for ocean (left panels) and land (right) grid points. Results are presented for different subensembles: resolutions (top panels), multi-physics (middle panels), and ocean boundary conditions (bottom panels). Errors are normalized by the median observed total precipitation across all regimes. The full and dashed red lines show the observational range using all five observations or only four data sets. For individual and total errors, relative error differences (in %) compared to the leftmost bar are shown. Refer to Sections 3.2 and 3.3 for details on the decomposition terms and the definition of errors.

decomposition provides a useful framework to evaluate models' precipitation that depends little on tunable parameters of the model, that provides physical insights about the sources of errors and that can be used to assess whether models are able to simulate the right amounts of rainfall for the right reasons. In addition, the proposed framework is well suited to assess whether models are fit-for-purpose for future projections because the background state quantities (circulation and moisture content) assessed will directly respond to future climate changes (Schmidt & Sherwood, 2015).

In our application of the ECIF decomposition to the precipitation associated to strong cyclones over the east coast of Australia, we found some clear improvements in the precipitation when running at convection-permitting resolutions with explicit representation of convection. We show that these improvements are not necessarily apparent when using common error metrics, such as mean biases or AEs, due to a strong compensation of errors. The added value of the convection-permitting simulations is thus made clear when the errors are separated into different components quantifying, in a meaningful way, different aspects of the simulation of precipitation. Our results also show substantial differences in the performance of various

model configurations related with choices of subgrid-scale schemes including cumulus, planetary boundary layer and MPS.

Our results suggest that the use of convection-permitting simulations (i.e., high resolution, explicit representation of convection) can bring important benefits. However, it is currently too costly to use convection-permitting resolutions for some climate applications and the decision about which specific model/configuration is to be used will depend not only on the models' performance but also on other considerations such as the models' complexity and the cost of performing ensembles of long simulations.

Further work is needed to elucidate some of these results including assessing in more detail the physical processes leading to differences between different configurations of the model. In addition, the application of this decomposition framework seems promising to assess future changes in precipitation in order to quantify dynamical and thermodynamical contributions to the total precipitation changes.

Data Availability Statement

The authors also thank Paola Petrelli and the Computational Modeling System team from the Australian Research Council's Centre of Excellence for Climate Extremes for providing access to observational data sets. Furthermore, the authors acknowledge the efforts of the European Centre for Medium-Range Weather Forecasts for providing the ERA-Interim and the ERA5 reanalyzes data. The authors thank Princeton Climate Analytics (PCA) for providing access to MSWEP data (www.gloh2o.org), the National Centers for Environmental Information for providing the CMORPH-CRT data and the Earth Observation Research Center of the Japan Aerospace Exploration Agency for providing the Global Rainfall Map in Near-Real-Time (GSMaP NRT).

Acknowledgments
This work was supported by Australian Research Council (ARC) grants DE170101191 and CE170100023. A. Di Luca was further funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant (RGPIN-2020-05631). D. Argüeso was financed by the EPICC project (PID2019-105253RJ-I00 MCI/AEI/FEDER,UE). The authors thank the National Center for Atmospheric Research and other participating institutions for making the WRF-ARW model available.

References

Ban, N., Schmidli, J., & Schär, C. (2014). Evaluation of the new convective-resolving regional climate modeling approach in decade-long simulations. *Journal of Geophysical Research: Atmospheres*, *119*(13), 7889–7907. https://doi.org/10.1002/2014JD021478

Bao, J., & Sherwood, S. C. (2019). The role of convective self-aggregation in extreme instantaneous versus daily precipitation. *Journal of Advances in Modeling Earth Systems*, *11*(1), 19–33. https://doi.org/10.1029/2018MS001503

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., Van Dijk, A. I., et al. (2019). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, *23*(1), 207–224. https://doi.org/10.5194/hess-23-207-2019

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., et al. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, *21*(12), 6201–6217. https://doi.org/10.5194/hess-21-6201-2017

Betts, A. K. (1986). A new convective adjustment scheme. Part I: Observational and theoretical basis. *Quarterly Journal of the Royal Meteorological Society*, *121*, 255–270. https://doi.org/10.1002/qj.49711247307

Bony, S., Dufresne, J. L., Le Treut, H., Morcrette, J. J., & Senior, C. (2004). On dynamic and thermodynamic components of cloud changes. *Climate Dynamics*, *22*(2–3), 71–86. https://doi.org/10.1007/s00382-003-0369-6

Brown, J. R., Jakob, C., & Haynes, J. M. (2010). An evaluation of rainfall frequency and intensity over the Australian region in a global climate model. *Journal of Climate*, *23*(24), 6504–6525. https://doi.org/10.1175/2010JCLI3571.1

Catto, J. L., Jakob, C., & Nicholls, N. (2015). Can the CMIP5 models represent winter frontal precipitation? *Geophysical Research Letters*, *42*, 8596–8604. https://doi.org/10.1002/2015GL066015

Catto, J. L., & Pfahl, S. (2013). The importance of fronts for extreme precipitation. *Journal of Geophysical Research: Atmospheres*, *118*(19), 10791–10810. https://doi.org/10.1002/jgrd.50852

Catto, J. L., Shaffrey, L. C., & Hodges, K. I. (2010). Can climate models capture the structure of extratropical cyclones? *Journal of Climate*, *23*(7), 1621–1635. https://doi.org/10.1175/2009JCLI3318.1

Cavicchia, L., Pepler, A., Dowdy, A., Evans, J., Di Luca, A., & Walsh, K. (2020). Future changes in the occurrence of hybrid cyclones: The added value of cyclone classification for the east australian low-pressure systems. *Geophysical Research Letters*, *47*(6), e2019GL085751. https://doi.org/10.1029/2019GL085751

Chambers, C. R., Brassington, G. B., Walsh, K., & Simmonds, I. (2015). Sensitivity of the distribution of thunderstorms to sea surface temperatures in four Australian east coast lows. *Meteorology and Atmospheric Physics*, *127*(5), 499–517. https://doi.org/10.1007/s00703-015-0382-4

Chambers, C. R., Chambers, C. R., Brassington, G. B., Simmonds, I., & Walsh, K. (2014). Precipitation changes due to the introduction of eddy-resolved sea surface temperatures into simulations of the "Pasha Bulker" Australian east coast low of June 2007. *Meteorology and Atmospheric Physics*, *125*(1–2), 1–15. https://doi.org/10.1007/s00703-014-0318-4

Chen, C.-T., & Knutson, T. (2008). On the verification and comparison of extreme rainfall indices from climate models. *Journal of Climate*, *21*(7), 1605–1621. https://doi.org/10.1175/2007JCLI1494.1

Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface-hydrology model with the penn state-ncar mm5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review*, *129*(4), 569–585. https://doi.org/10.1175/1520-0493(2001)129<0569:caalsh>2.0.co;2

Collins, W. D., Rash, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., & Briegleb, B. (2004). *Description of the NCAR community atmosphere model (CAM 3.0)*. (Tech. Rep.). NCAR/TN-464+STR. NCAR Technical Note.

Dai, A., Rasmussen, R. M., Liu, C., Ikeda, K., & Prein, A. F. (2017). A new mechanism for warm-season precipitation response to global warming based on convection-permitting simulations. *Climate Dynamics*, *55*(1), 343–368. https://doi.org/10.1007/s00382-017-3787-6

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Di Luca, A., Argüeso, D., Evans, J. P., de Elía, R., & Laprise, R. (2016). Quantifying the overall added value of dynamical downscaling and the contribution from different spatial scales. *Journal of Geophysical Research: Atmospheres*, *121*(4), 1575–1590. https://doi.org/10.1002/2015JD024009

Di Luca, A., Argüeso, D., Jourdain, N., & Evans, J. P. (2021). *High-resolution modeling of extreme storms over the east coast of Australia v1.0, NCI national research data collection*. https://geonetwork.nci.org.au/geonetwork/srv/eng/catalog.search#/metadata/f3907_9375_5188_8823

Di Luca, A., Evans, J. J. P., Pepler, A. S. A., Alexander, L. V. L., & Argüeso, D. (2016). Evaluating the representation of Australian East Coast Lows in a regional climate model ensemble. *Journal of Southern Hemisphere Earth System Science*, *66*(2), 108–124. https://doi.org/10.22499/3.6602.003

Di Luca, A., Pitman, A. J., & de Elía, R. (2020). Decomposing temperature extremes errors in CMIP5 and CMIP6 models. *Geophysical Research Letters*, *47*(14), 1–10. https://doi.org/10.1029/2020GL088031

Dowdy, A. J., & Catto, J. L. (2017). Extreme weather caused by concurrent cyclone, front and thunderstorm occurrences. *Scientific Reports*, *7*, 1–8. https://doi.org/10.1038/srep40359

Dowdy, A. J., Pepler, A., Di Luca, A., Cavicchia, L., Mills, G., Evans, J. P., et al. (2019). Review of Australian east coast low pressure systems and associated extremes. *Climate Dynamics*, *53*(7), 4887–4910. https://doi.org/10.1007/s00382-019-04836-8

Dudhia, J. (1989). Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, *46*, 3077–3107. https://doi.org/10.1175/1520-0469(1989)046<3077:nsocod>2.0.co;2

Eyring, V., Gleckler, P. J., Heinze, C., Stouffer, R. J., Taylor, K. E., Balaji, V., et al. (2016). Toward improved and more routine Earth system model evaluation in CMIP. *Earth System Dynamics*, *7*(4), 813–830. https://doi.org/10.5194/esd-7-813-2016

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., et al. (2015). The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Scientific Data*, *2*(1), 150066. https://doi.org/10.1038/sdata.2015.66

Gilmore, J. B., Evans, J. P., Sherwood, S. C., Ekström, M., & Ji, F. (2016). Extreme precipitation in WRF during the Newcastle east coast low of 2007. *Theoretical and Applied Climatology*, *125*(3), 809–827. https://doi.org/10.1007/s00704-015-1551-6

Giorgi, F., & Bi, X. (2000). A study of internal variability of a regional climate model. *Journal of Geophysical Research*, *105*(D24), 29503–29521. https://doi.org/10.1029/2000JD900269

Govekar, P. D., Jakob, C., & Catto, J. (2014). The relationship between clouds and dynamics in Southern Hemisphere extratropical cyclones in the real world and a climate model. *Journal of Geophysical Research: Atmospheres*, *119*, 6609–6628. https://doi.org/10.1002/2013JD020699

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Holloway, C. E., & Neelin, D. J. (2009). Moisture vertical structure, column water vapor, and tropical deep convection. *Journal of the Atmospheric Sciences*, *66*(6), 1665–1683. https://doi.org/10.1175/2008JAS2806.1

Hong, S. Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*, 2318–2341. https://doi.org/10.1175/mwr3199.1

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., & Tan, J. (2019). *Integrated multi-satellite retrievals for GPM (IMERG) technical documentation*. NASA. (Tech. Rep.). Retrieved from https://pmm.nasa.gov/sites/default/files/document_files/IMERG_doc_190909.pdf

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., et al. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, *8*(1), 38–55. https://doi.org/10.1175/JHM560.1

Huffman, G. J., Stocker, E., Bolvin, D., Nelkin, E., & Jackson, T. (2019). *GPM IMERG final precipitation l3 half hourly 0.1 degree x 0.1 degree v06*. https://doi.org/10.5067/GPM/IMERG/3B-HH/06

Janjic, Z. I. (1994). The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer and turbulence closure schemes. *Monthly Weather Review*, *122*, 927–945.

Janjic, Z. I. (2000). Comments on "Development and evaluation of a convection scheme for use in climate models". *Journal of the Atmospheric Sciences*, *57*, 3686.

Jones, D. A., Wang, W., & Fawcett, R. (2009). High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal*, *58*(4), 233–248. https://doi.org/10.22499/2.5804.003

Joyce, R. J., Janowiak, J. E., Arkin, P. A., & Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, *5*(3), 487–503. https://doi.org/10.1175/1525-7541(2004)005<0487:camtpg>2.0.co;2

Kain, J. S. (2004). The Kain-Fritsch convective parameterization: An update. *Journal of Applied Meteorology*, *43*, 170–181. https://doi.org/10.1175/1520-0450(2004)043<0170:tkcpau>2.0.co;2

Kain, J. S., & Fritsch, J. M. (1993). Convective parameterization for mesoscale models: The Kain-Fritsch scheme. In K. A. Emanuel, & D. J. Raymond (Eds.), *The representation of cumulus convection in numerical models*. American Meteorological Society. https://doi.org/10.1007/978-1-935704-13-3_16

Kidd, C., & Levizzani, V. (2011). Status of satellite precipitation retrievals. *Hydrology and Earth System Sciences*, *15*(4), 1109–1116. https://doi.org/10.5194/hess-15-1109-2011

Knist, S., Goergen, K., & Simmer, C. (2018). Evaluation and projected changes of precipitation statistics in convection-permitting WRF climate simulations over Central Europe. *Climate Dynamics*, *0*(0), 1–341. https://doi.org/10.1007/s00382-018-4147-x

Kubota, T., Aonashi, K., Ushio, T., Shige, S., Takayabu, Y. N., Kachi, M., et al. (2020). Global satellite mapping of precipitation (GSMaP) products in the GPM ERA. In V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, & F. J. Turk (Eds.), *Satellite precipitation measurement* (Vol. 1, pp. 355–373). Springer International Publishing. https://doi.org/10.1007/978-3-030-24568-9_20

Kubota, T., Shige, S., Hashizume, H., Aonashi, K., Takahashi, N., Seto, S., et al. (2007). Global precipitation map using satellite-borne microwave radiometers by the GSMaP project: Production and validation. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(7), 2259–2275. https://doi.org/10.1109/tgrs.2007.895337

Lim, K. S. S., & Hong, S. Y. (2010). Development of an effective double-moment cloud microphysics scheme with prognostic Cloud Condensation Nuclei (CCN) for weather and climate models. *Monthly Weather Review*, *138*, 1587–1612. https://doi.org/10.1175/2009mwr2968.1

Louis, S., Couriel, E., Lewis, G., Glatz, M., Kulmar, M., Golding, G., & Hanslow, D. (2016). NSW east coast low event – 3 to 7 June 2016. In *25th NSW Coastal Conference Coffs Harbor, NSW* (pp. 1–16).

Mega, T., Ushio, T., Takahiro, M., Kubota, T., Kachi, M., & Oki, R. (2019). Gauge-adjusted global satellite mapping of precipitation. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(4), 1928–1935. https://doi.org/10.1109/tgrs.2018.2870199

Mills, G. A., Webb, R., Davidson, N. E., Kepert, J., Seed, A., & Abbs, D. (2010). *The pasha bulker east coast low of 8 june 2007*. (Tech. Rep.). Centre for Australia Weather and Climate Research Tech. Rep. 23.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the long-wave. *Journal of Geophysical Research*, *102*(D14), 16663–16682. https://doi.org/10.1029/97jd00237

Oke, P. R., Sakov, P., Cahill, M. L., Dunn, J. R., Fiedler, R., Griffin, D. A., et al. (2013). Toward a dynamically balanced eddy-resolving ocean reanalysis: BRAN3. *Ocean Modelling*, *67*, 52–70. https://doi.org/10.1016/j.ocemod.2013.03.008

Palmer, T. N. (2016). A personal perspective on modeling the climate system. *Proceedings of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *472*, 20150772. https://doi.org/10.1098/rspa.2015.0772

Pepler, A. S., Di Luca, A., & Evans, J. P. (2017). Independently assessing the representation of midlatitude cyclones in high-resolution reanalyzes using satellite observed winds. *International Journal of Climatology*, *38*, 1314–1327. https://doi.org/10.1002/joc.5245

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., et al. (2017). The weather research and forecasting model: Overview, system efforts, and future directions. *Bulletin of the American Meteorological Society*, *98*(8), 1717–1737. https://doi.org/10.1175/BAMS-D-15-00308.1

Prein, A. F., Gobiet, A., Truhetz, H., Keuler, K., Goergen, K., Teichmann, C., et al. (2016). Precipitation in the EURO-CORDEX 0.11° and 0.44° simulations: High resolution, high benefits? *Climate Dynamics*, *46*(1–2), 383–412. https://doi.org/10.1007/s00382-015-2589-y

Quinting, J. F., Catto, J. L., & Reeder, M. J. (2019). Synoptic climatology of hybrid cyclones in the Australian region. *Quarterly Journal of the Royal Meteorological Society*, *145*(718), 288–302. https://doi.org/10.1002/qj.3431

Rajczak, J., & Schär, C. (2017). Projections of future precipitation extremes over Europe: A multimodel assessment of climate simulations. *Journal of Geophysical Research: Atmospheres*, *122*(20), 10773–10800. https://doi.org/10.1002/2017JD027176

Sahany, S., Neelin, J. D., Hales, K., & Neale, R. B. (2014). Deep convective transition characteristics in the community climate system model and changes under global warming. *Journal of Climate*, *27*(24), 9214–9232. https://doi.org/10.1175/JCLI-D-13-00747.1

Schmidt, G. A., & Sherwood, S. (2015). A practical philosophy of complex climate modeling. *European Journal for Philosophy of Science*, *5*(2), 149–169. https://doi.org/10.1007/s13194-014-0102-9

Separovic, L., de Elía, R., & Laprise, R. (2011). Impact of spectral nudging and domain size in studies of RCM response to parameter modification. *Climate Dynamics*, *38*(7–8), 1325–1343. https://doi.org/10.1007/s00382-011-1072-7

Singh, M. S., & O'Gorman, P. A. (2014). Influence of microphysics on the scaling of precipitation extremes with temperature. *Geophysical Research Letters*, *41*(16), 6037–6044. https://doi.org/10.1002/2014GL061222

Skamarock, W., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M., & Powers, J. G. (2008). *A description of the advanced research WRF version 3* (Tech. Rep. No. June). DTIC Document. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord{\&}metadataPrefix=html{\&}identifier=ADA487419

Skamarock, W. C. (2004). Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly Weather Review*, *132*(12), 3019–3032. https://doi.org/10.1175/mwr2830.1

Stensrud, D. J. (2007). *Parameterization schemes: Keys to understanding numerical weather prediction models*. Cambridge University Press. https://doi.org/10.1017/CBO9780511812590

Sun, Y., Solomon, S., Dai, A., & Portmann, R. W. (2006). How often does it rain? *Journal of Climate*, *19*, 916–934. https://doi.org/10.1175/JCLI3672.1.1996

TRMM. (2011). *TRMM (TMPA) rainfall estimate l3 3 hour 0.25 degree x 0.25 degree v7*. https://doi.org/10.5067/TRMM/TMPA/3H/7

Williams, K. D., & Tselioudis, G. (2007). GCM intercomparison of global cloud regimes: Present-day evaluation and climate change response. *Climate Dynamics*, *29*(2–3), 231–250. https://doi.org/10.1007/s00382-007-0232-2

Xie, P., Joyce, R., Wu, S., Yoo, S.-H., Yarosh, Y., Sun, F., & Lin, R. (2017). Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. *Journal of Hydrometeorology*, *18*(6), 1617–1641. https://doi.org/10.1175/JHM-D-16-0168.1

Xie, P., Joyce, R., Wu, S., Yoo, S.-H., Yarosh, Y., Sun, F., & Lin, R. (2019). *NOAA climate data record (CDR) of CPC morphing technique (CMORPH) high resolution global precipitation estimates, version 1, hourly-0.25* (Tech. Rep.). NOAA National Centers for Environmental Information. https://doi.org/10.25921/w9va-q159

Zappa, G., Hawcroft, M. K., Shaffrey, L., Black, E., & Brayshaw, D. J. (2015). Extratropical cyclones and the projected decline of winter Mediterranean precipitation in the CMIP5 models. *Climate Dynamics*. https://doi.org/10.1007/s00382-014-2426-8

Zappa, G., Shaffrey, L. C., & Hodges, K. I. (2013). The ability of CMIP5 models to simulate North Atlantic extratropical cyclones*. *Journal of Climate*, *26*(15), 5379–5396. https://doi.org/10.1175/JCLI-D-12-00501.1