

## RESEARCH ARTICLE

10.1029/2017JD028192

## Key Points:

- Biases in model clouds and rain arise from the partitioning of water between hydrometeor species
- A small subset of parameterization changes can reproduce the variations between microphysics schemes
- The number of prognostic variables has less effect than the choice of subgrid-scale closures

## Correspondence to:

K. Furtado,  
kalli.furtado@lasg.iap.ac.cn

## Citation:

Furtado, K., Field, P. R., Luo, Y., Liu, X., Guo, Z., Zhou, T., et al. (2018). Cloud microphysical factors affecting simulations of deep convection during the presummer rainy season in Southern China. *Journal of Geophysical Research: Atmospheres*, 123, 10,477–10,505. <https://doi.org/10.1029/2017JD028192>

Received 13 DEC 2017

Accepted 6 JUL 2018

Accepted article online 18 JUL 2018

Published online 18 SEP 2018

## Cloud Microphysical Factors Affecting Simulations of Deep Convection During the Presummer Rainy Season in Southern China

Kalli Furtado<sup>1</sup> , Paul R. Field<sup>1</sup> , Yali Luo<sup>2</sup> , Xi Liu<sup>2</sup>, Zhun Guo<sup>3</sup>, Tianjun Zhou<sup>3</sup> , Benjamin J. Shipway<sup>1</sup>, Adrian A. Hill<sup>2</sup>, and Jonathan M. Wilkinson<sup>1</sup> 

<sup>1</sup>Met Office, Exeter, UK, <sup>2</sup>Chinese Academy of Meteorological Sciences, Beijing, China, <sup>3</sup>Institute of Atmospheric Physics, Beijing, China

**Abstract** The sensitivity of subtropical deep convection to the parameterization of cloud microphysics is elucidated through high-resolution modeling of extreme presummer rainfall over southern China. An ensemble of physics configuration experiments is used to identify several drivers of model errors in comparison to radar observations from the South China Monsoon Rainfall Experiment (SCMREX) and remotely sensed estimates of cloud, precipitation, and radiation from satellites in the A-train constellation. The benefits of increasing the number of prognostic variables in the microphysics scheme is assessed, relative to the effects of the parameterization of cloud microphysical properties and cloud fraction diagnosis. By matching individual parameterizations between the microphysical configurations, it is shown that a small subset of the parameterization changes can reproduce most of the dependence of model performance on physics configuration. In particular, biases that are due to the low-level clouds and rain are strongly influenced by cloud fraction diagnosis and raindrop size distribution, whereas variations in the effects of high clouds are strongly influenced by differences in the parameterization of ice crystal sedimentation. Hence, for the case studied here, these parameterizations give more insight into the causes of variability in model performance than does the number of model prognostics per se.

### 1. Introduction

Accurate modeling of organized deep convection is important for predicting many phenomena in the tropical and subtropical atmosphere, such as monsoons, tropical cyclones, and the diurnal cycle (Holloway et al., 2014; Waliser et al., 2012). On climate time scales the representation of convection in models is often the most important factor affecting regional- and planetary-scale biases in clouds, radiation, and the cycling of water within the Earth system (Stevens & Bony, 2013). Climate sensitivity (Sherwood et al., 2014) and climate impacts (e.g., flood-inducing extremes of precipitation; Doswell et al., 1996) are often also influenced by convective cloud systems, and therefore, deficiencies in the simulation of convection can have a limiting effect on the veracity of regional forecasts and on the development of effective climate services.

The presummer rainy season (April–June) in South China is an early stage of the summer monsoon in East Asia. It peaks in May, with the onset of the monsoon in the South China Sea and ends in June when the monsoon rainband shifts north to the Yangtze River valley (Ding & Chan, 2005; P. Li et al., 2017). During this period, the region experiences approximately half of its annual accumulated rainfall, due to the frequent occurrence of extreme rainfall associated with organized mesoscale convective systems (MCSs; with horizontal extents of 200–2,000 km and typical lifetimes of about 18 hr), leading to severe flooding that endangers lives and livelihoods (Luo et al., 2017). There has been steady progress in understanding the mechanisms leading to formation of these heavy-rain-producing MCSs (S. S. Huang, 1986; Luo et al., 2017; Zhang et al., 2011; Zhou, 2003), but the skill of regional precipitation forecasts remains low (particularly for extreme rainfall; L. Huang & Luo, 2017). This is in part due to a lack of understanding of the microphysical and dynamical drivers that are operating and how well these are represented by numerical weather models (Luo et al., 2017).

Moreover, MCSs are an important hydrological component of the East Asian Summer Monsoon and are a frequent phenomena in China in general. They often develop as part of mesoscale and synoptic-scale cloud systems, particularly on the warm-air (southern) sides of fronts and shear lines, due to the favorable conditions

created by a combination of warm advection by lower tropospheric jets (southwesterlies) and diffuence in the upper troposphere (see, e.g., Ding & Chan, 2005, and references therein). As such MCSs are a key ingredient of the presummer rainy season in South China, as well as the Meiyu season in the Yangtze and Huaihe River basins (also known as the Baiyu season in Japan and Changma season in Korea).

Because evolution of deep convective clouds is influenced by small-scale microphysical processes, how these processes are represented in models affects the realism of simulated clouds. The relative complexity of cloud microphysics schemes in atmospheric models is an attempt to achieve a sufficient level of physical realism in the representation of cloud processes. This complexity arises primarily from two interdependent sources: the number of microphysical processes included in a scheme, and the level of sophistication, or physical detail, to which those processes are modeled. A relatively simple scheme may use a single prognostic variable to represent each of a small number of cloud species (e.g., cloud droplets, rain, and ice), whereas a more complex scheme will typically involve finer differentiation of cloud species (graupel, hail, cloud ice, and snow) and employ more prognostics for each species. The greater differentiation of species leads to more interspecies interactions and the increased number of prognostics leads to greater sophistication in the representation of these processes. This increased complexity raises the computational cost, but it is not obvious what the effects of this additional complexity are, or where the balance between costs and benefits lies.

Microphysics schemes can also differ in terms of the parameterizations used for the same physical process. For example, schemes rely on a large number of assumed hydrometeor properties, which are needed to provide a closed description of the processes involved, and these properties can vary widely between schemes. Closure relations include particle size distributions and parameterized mass-diameter and fallspeed-diameter relationships. When comparing microphysics schemes in order to understand a physical process, or optimizing a given scheme for a particular application (e.g., weather forecasting), it is of interest to understand which factors are the main drivers of intermodel spread and model performance.

In situations where model performance is influenced by the complexity of the microphysics scheme, it is pertinent to ask the following: (a) what are the mechanisms by which microphysical complexity affects the model predictions? Or, in other words, which processes are most sensitive to microphysical complexity and why? And (b) to what extent can a less complex scheme be *tuned* so as to emulate the effects of greater complexity? A route to answering these questions is to identify how complexity affects individual microphysical processes and the implications that these effects have for simulated cloud properties. This approach was employed by Morrison et al. (2009) to investigate the effects of microphysics scheme complexity on simulations of idealized squall lines, and by Van Weverberg et al. (2013) and Varble et al. (2014) in modeling studies of observed convective systems in the tropics. These studies identified a range of predicted properties which differed between single- and double-moment microphysics schemes. For example, Van Weverberg et al. (2013) showed that the mass of ice in the upper troposphere varied between schemes with different levels of complexity, and Varble et al. (2014) showed that rain water contents were highest in single-moment schemes. The results of Morrison et al. (2009) suggested that there are intrinsic advantages to higher-complexity (multi-moment) schemes, because they are able to represent a range of rain-producing processes within the same computational domain.

In answering questions (a) and (b), above, it is useful to bear in mind that many cloud properties are influenced by only a small number of microphysical processes. Two simple examples of this are as follows: the condensed water contents of midlatitude frontal clouds (Furtado et al., 2015) and tropical deep convection (Van Weverberg et al., 2013), which depend strongly on the parameterization of the hydrometeor sedimentation flux; rain water contents and rainfall rates, which are sensitive to the parameterization of raindrop number concentration (Morrison et al., 2009; Varble et al., 2014). Recent work by Eidhammer et al. (2017) comparing traditional *speciated* ice microphysics (Gettelman & Morrison, 2015) to a *single-category-variable-habit* approach (Morrison & Milbrandt, 2015), has shown that the differences in cloud radiative forcing between the schemes were due to differences in the parameterization of ice particle mean fallspeeds. It is not likely that all microphysics schemes will respond in the same way to all changes—there are structural differences between schemes which may preclude this possibility. However, because all schemes have been designed to incorporate the same physics (the physics of clouds), we may expect commonalities between them in terms of how and why they perform in a given way for a given type of cloud. The above examples suggest that sensitivity to fallspeed and number concentration is an area where we expect such commonalities to exist. Hence, these parameters will be a focus of this study. Indeed, there may be simple physical reasons why sedimentation

has been identified as an important factor in many models. For example, Furtado et al. (2015) showed that mass conservation of water places an analytical constraint on how clouds respond to changes in hydrometeor fallspeed, because (in quasi steady state conditions) the precipitation flux must balance the upward flux of water vapor.

An additional source of complexity in numerical models is the parameterization of condensation of water vapor to form liquid cloud droplets. Because the time scale for condensation is extremely short, compared to a typical model time step, the initial formation of liquid cloud can be modeled by diagnostically dividing the total mass of vapor *plus* liquid water between condensed and gaseous phases. Because the size of a grid-box in a model can exceed the spatial scale of heterogeneities in atmospheric water content, models often accomplish this partitioning by making assumptions about the statistical properties of humidity on subgrid scales: so-called *cloud macrophysics* (Sommeria & Deardorf, 1977). There is considerable uncertainty about these assumptions (particularly in the presence of ice crystals), and the relative importance of cloud macrophysics, compared to the subsequent microphysical evolution of in-cloud properties, needs to be assessed according to the cloud regime and application under consideration.

In this study, we use a case study of organized deep convection over South China to identify a set of major drivers of model errors in the properties of deep convective clouds and rainfall and, in particular, how these errors are affected by the level of complexity inherent in the cloud microphysics and macrophysics schemes.

## 2. Description of the Model Evaluation Framework and SCMREX Case Studies

The atmospheric model used in this study is a high-resolution (convection permitting) local area model (LAM) configuration of the Met Office Unified Model, nested inside a coarser-resolution global simulation performed with the Met Office Global Atmosphere 6.1 configuration (Walters et al., 2017). Lateral-boundary conditions for the LAM are applied hourly in a rim of width 9 grid lengths around the edge of the model domain. The high-resolution domain has an angular grid spacing of  $0.04^\circ$ , which corresponds to an approximate grid length of 4 km in the model's rotated-pole coordinate system. The vertical level set has 80 terrain-following levels with a model top at 35.8 km and a lowest model level at 100 m. Boundary conditions are derived hourly from the global driving simulation, which has a grid spacing of  $0.2^\circ$ . The dynamical core of both models is nonhydrostatic and uses semi-Lagrangian advection. The driving model uses a mass-flux-based convection parameterization, but this is turned off in the high-resolution LAM. The boundary layer parameterization operates on all model levels and is based on the nonlocal scheme described by Lock et al. (2000), with modifications to make the scheme more applicable to convective *gray zone* scales. The radiation scheme in the model is based on Edwards and Slingo (1996) and calculates a two-stream approximation to a spectrally resolved radiance field, subject to scattering and absorption by clouds, aerosols, and atmospheric gases. The optical properties of ice crystals in the scheme use the habit-mixture approach described by Baran et al. (2014). In general, the model physics in the LAM is similar to a configuration used operationally by Met Office for weather forecasting applications over the United Kingdom.

As a framework for varying the complexity of the microphysics within the Unified Model, we use the Cloud and AeroSol Interacting Microphysics (CASIM) scheme. The scheme allows for a range of complexity, ranging from all-single-moment (CASIM 1M) to a fully double-moment scheme (CASIM 2M; Grosvenor et al., 2017; Miltenberger et al., 2018). Triple-moment configurations are also permitted, but will not be investigated in this paper. The default CASIM configurations have five hydrometeor species: cloud droplets, raindrops, cloud ice crystals, snow aggregates, and graupel. The following structural scheme-elements are of relevance to understanding the results of the sensitivity tests. Warm-cloud microphysics is based on autoconversion of cloud droplets to rain and subsequent collisional coalescence of the slowly settling cloud droplets with more rapidly sedimenting raindrops. Raindrops can exchange mass with the vapor phase by evaporation (dependent on the mean supersaturation in a grid box), and the time scale for this process is modeled explicitly and depends on the first moment of the raindrop size distribution (DSD). By contrast, cloud droplets are assumed to condense out instantaneously (relative to the duration of a model time step). Hence, the mass of liquid cloud is determined by a *saturation adjustment*, that is, a diagnostic partitioning of total (liquid *plus* vapor) water between gaseous and condensed phases, based on Smith, (1990; as implemented by Grosvenor et al., 2017), which depends on a parameterized threshold for cloud formation (the critical relative humidity,  $RH_c$ ). Small ice crystals can form due to freezing of liquid water droplets by either homogeneous nucleation (at a temperature

**Table 1**  
*Microphysics Configurations Intercompared in This Study*

Name	Iris g	PSD type	Ice shape	DSD closure	Cloud
2M	22222	gamma	sphere	prognostic, $\mu_r = 2.5$	on/off
2M CS	22222	gamma	sphere	prognostic, $\mu_r = 2.5$	$RH_c$
1M	11111	gamma	sphere	$N_r = 5 \cdot 10^6$ , $\mu_r = 2.5$	on/off
1M CS	11111	gamma	sphere	$N_r = 5 \cdot 10^6$ , $\mu_r = 2.5$	$RH_c$
1M Nr	11111	gamma	sphere	$N_r = 2.5 \cdot 10^4$ , $\mu_r = 2.5$	on/off
1M Nr SI	11–11	mixed <sup>a</sup>	nonsph. <sup>b</sup>	$N_r = 2.5 \cdot 10^6$ , $\mu_r = 2.5$	on/off
WB	11–11	generic	nonsph.	intercept <sup>c</sup> , $\mu_r = 0$	on/off
WB CS	11–11	generic	nonsph.	intercept, $\mu_r = 0$	$RH_c$

<sup>a</sup>PSD is generic for sedimentation, gamma for all other processes. <sup>b</sup>Nonspherical,  $M \propto D^2$  (disk-like prefractal), for all processes. <sup>c</sup> $N_r = n_d \lambda^{1-n_b}$ .

threshold of  $-40^\circ\text{C}$ ) or by heterogeneous freezing using a parameterization of immersion freezing. Once ice crystals are present, they can be transferred to the snow category by autoconversion and subsequent collision accretion. Graupel is formed by collisions between ice crystals and snow aggregates with cloud droplets or raindrops. Ice, snow, and graupel can all sediment, grow (and sublimate) by vapor deposition, and be converted to cloud or rain by melting at temperatures warmer than  $0^\circ\text{C}$ . The above process rates depend on the particle size distributions,  $f_s(D)$ , and the sets of fixed microphysical properties,  $\mathcal{P}_s$ , which are assumed for the various species,  $s \in \{\text{cloud, rain, ...}\}$ , of hydrometeor. Among the microphysical properties belonging to  $\mathcal{P}_s$ , of relevance here are the parameters,  $a_s$ ,  $b_s$ ,  $c_s$ , and  $d_s$ , in the mass-diameter ( $M_s = a_s D^{b_s}$ ) and fallspeed-diameter ( $V_s \propto c_s D^{d_s}$ ) parameterizations for each species. These quantities enter into the usual single- and double-moment closure relations to give closed expressions of the process rates in terms of the prognostic variables. For all five cloud species, the particles are assumed to be spherical in shape (or, at least, to have masses that are proportional to the cube of their linear dimensions; corresponding to  $b_s = 3$  for all  $s$ ). The effective densities,  $6a_s/\pi$ , of cloud droplets and raindrops are equal to the material density of liquid water ( $1,000 \text{ kg/m}^3$ ), whereas the nature of solid phase cloud particles is incorporated by using effective densities of  $200 \text{ kg/m}^3$  and  $100 \text{ kg/m}^3$  for ice crystals and snow aggregates, respectively. For a double-moment species, the process rates are closed in terms of the prognostic *mass mixing ratio*,  $q_s$  ( $\text{kg/kg}$ ), and *number*,  $N_s$  ( $\text{kg}^{-1}$ ), of the species. For a single-moment species, the sole prognostic variable is  $q_s$  and the number concentration is a species-dependent constant (see Table 1). In both cases, the size distributions of all the species are assumed to be Gamma distributions:  $f_s(D) = N_s \Gamma(\mu_s + 1) D_s^{\mu_s} \exp(-\lambda_s D) / \lambda_s^{\mu_s + 1}$ , where  $\mu_s$  is an  $s$ -dependent constant and  $\lambda_s = \Lambda(q_s, N_s, \mathcal{P}_s)$  is the function of the prognostics and microphysical parameters that is determined from the single- or double-moment closure relations.

The CASIM simulations are compared to simulations using the Wilson and Ballard (1999, WB) microphysics scheme that is currently used operationally in the Unified Model for weather forecasting and climate projection. The WB scheme has four prognostic hydrometeor species: cloud, rain, a single (*generic*) ice-snow category, and graupel. Structurally, the WB scheme is similar to the CASIM M1, except for the following noteworthy differences. First, the absence of a separate ice crystal category, in addition to the generic ice-snow species, means that ice-to-snow autoconversion and snow-ice collisions are not modeled. Second, the treatment of graupel is slightly different and follows the framework described in Wilkinson (2017). The majority of differences between WB and CASIM, therefore, involve either different choices for the parameterizations of specific physical processes (e.g., autoconversion) or different choices of microphysical properties (e.g., size distributions, mass-diameter relations). For rain and graupel the size distributions are Gamma distributions, as in the CASIM schemes, but the diagnostic number concentrations of these species are parameterized to vary with mean particle size:  $N_s = n_{0s} \lambda^{1+\mu_s-n_{1s}} / \Gamma(1 + \mu_s)$ , where  $n_{0s}$  and  $n_{1s}$  are constants and  $s \in \{\text{rain, graupel}\}$  (cf., CAISM M1, for which  $N_r$  and  $N_g$  are constants). For rain, the values of these parameters were derived by Abel and Boutle (2012) from aircraft measurements, with the aim of representing the observed relationship between the mass and sedimentation flux of raindrops. A further difference between the CASIM and WB schemes is that the size distribution for the generic ice-snow category in the WB scheme uses the moment estimation parameterization developed by Field et al. (2007), together with a mass-diameter relationship that is characteristic of nonspherical ice particles (Cotton et al., 2012). The fallspeed-diameter relation for ice also

differs from the CASIM scheme and uses the parameterization from Furtado et al. (2015) (see model configuration *W*, therein). These differences imply that, for the same ice water content, there may be significant differences in the ice microphysical process rates between the CASIM and WB simulations.

Both the CASIM and WB schemes allow the complexity of the diagnostic cloud scheme to be varied: a simple *on-off* cloud scheme, for which only subgrid cloud fractions of one (fully overcast) or zero (no cloud) are allowed, can be compared to a more complex scheme that accounts for subgrid-scale moisture variability. In both schemes, the subgrid volume fraction of liquid cloud,  $C_l$ , is a diagnostic function of the gridbox mean variables (temperature, water content, etc.) and depends on a vertical profile of adjustable parameters,  $RH_c$ , as described by Smith (1990) and Grosvenor et al. (2017). A purely *on-off* cloud fraction scheme corresponds to the choice  $RH_c(z_h) = 1$ , for all values of the model's terrain-following vertical coordinate,  $z_h$ . Subsequently,  $RH_c$  will be referred to as the *critical relative humidity* (for the formation of liquid cloud).

The function  $C_l$  has a natural interpretation as the zeroth moment of a probability density function (PDF) for subgrid-scale fluctuations in water content and temperature (see, e.g., Sommeria & Deardorf, 1977; Smith, 1990). For the CASIM and WB schemes,  $C_l$  is consistent with a symmetric triangular PDF with half width equal to  $RH_c$ . The cloud water content,  $q_l$ , is determined from a truncated first moment of the PDF (Grosvenor et al., 2017; Smith, 1990). To identify the relative importance of cloud fraction diagnosis for model biases, simulations with  $RH_c = 1$  will be compared to simulations with a vertical profile of values of  $RH_c$ . In terms of the subgrid scale PDFs, a value of  $RH_c = 1$  corresponds to a very sharply peaked (narrow) distribution, for which the gridbox mean water content must attain water saturation before any liquid cloud can be diagnosed in a gridbox. It should be noted that, in this paper, where a value of  $RH_c \neq 1$  is used, only liquid clouds are directly effected, that is, in the experiments considered here  $RH_c$  has no effect on the diagnosed ice cloud fractions, which are always assumed to be *on/off*.

### 2.1. Details of the Model Sensitivity Tests

Simulations are performed with three *baseline* model configurations and with a set of sensitivity tests designed to investigate the causes of the main differences between baseline simulations (Table 1). As baseline simulations we use CASIM all-double-moment (thereafter 2M), CASIM all-single-moment (1M) and the default configuration of the Wilson-Ballard scheme (WB). In all the baseline configurations the critical relative humidity for cloud formation,  $RH_c$ , is equal to one on all model levels. The additional sensitivity tests are as follows:

#### 2.1.1. CASIM All Single Moment With Modified Raindrop Number Concentration (1M Nr)

In the CASIM single-moment scheme, the default value of the diagnostic number concentration,  $N_r$ , of raindrops is  $5 \times 10^6 \text{ kg}^{-1}$ . To investigate the effects of this parameter, we will compare to a sensitivity test for which  $N_r$  is reduced to  $2.5 \times 10^4 \text{ kg}^{-1}$ . It will be shown below that the lower value is typical of the number concentration that arises in the 2M simulation. Hence, reducing  $N_r$  gives diagnostic number concentrations that resemble those in the 2M simulation.

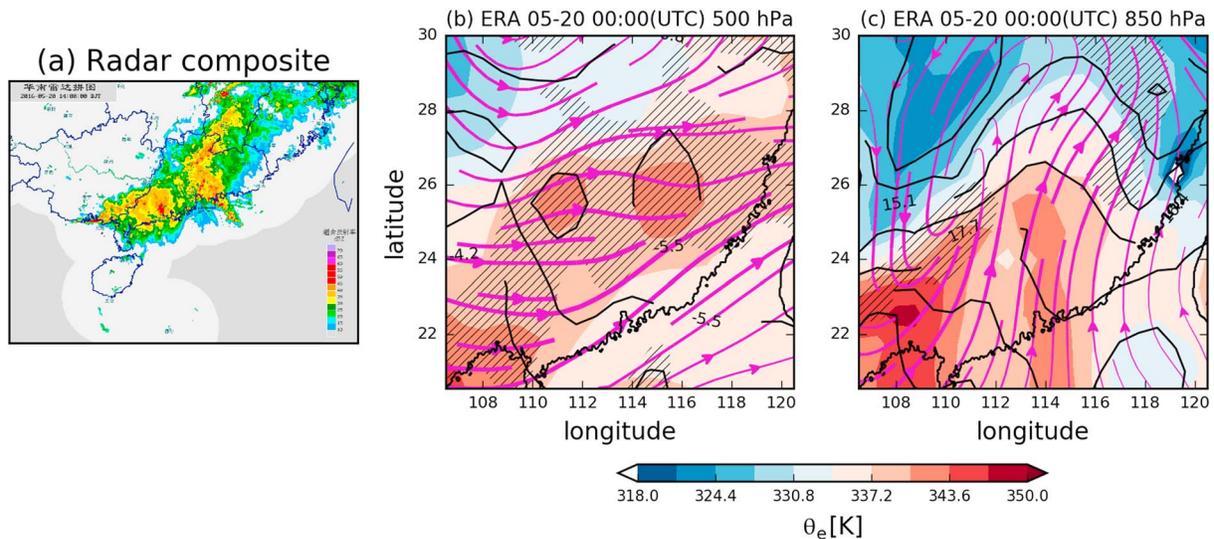
#### 2.1.2. CASIM All Single Moment With Modified Ice Microphysics (1M SI)

The default configurations of CASIM have separate prognostic variables for small ice crystals and larger snow aggregates, particles of both of which are assumed to be spherical and gamma distributed in their linear sizes. By contrast, the Wilson-Ballard scheme uses a single prognostic variable for a combined (*generic*) snow-ice species, which is assumed to be nonspherical in shape and is distributed according to the bimodal size distribution derived by Field et al. (2007). To understand the role of these differences, we modify the CASIM 1M scheme by (a) removing the snow aggregate category, so that there is only one hydrometeor species for ice-snow; (b) using the WB mass-diameter relation for the particles of the single ice-snow species; (c) modifying the parameterization of mass-weighted mean fallspeed so that it uses the same PSD and the fallspeed-diameter relations as the WB scheme. Note that the PSD and fallspeeds are modified *only* in parameterization of mean sedimentation flux; other microphysical processes are not affected by the changes. This is done to identify the important role of sedimentation in affecting intermodel differences.

To consistently assess differences in radar reflectivity between the WB and CASIM schemes, the radar reflectivity calculations for the 1M SI configuration uses the Field et al. (2007) PSD.

#### 2.1.3. CASIM 1M Nr SI

A combination of the modified raindrop number and modified ice microphysics experiments.



**Figure 1.** The synoptic conditions associated with the SR1 heavy rainfall event at 00 UTC 20 May 2016. (a) Composite radar imagery (maximum reflectivity in column), from Chinese Meteorological Administration’s network of operational weather radars over south China. (b, c) The ERA Interim reanalysis for 00 UTC 20 May, at 500 hPa and 850 hPa: equivalent potential temperature ( $\theta_e$ , red-blue colors); regions of negative pressure-velocity ( $\omega$  at 500 hPa, hatches) and regions of negative horizontal convergence ( $\nabla_h \cdot \mathbf{u}$  at 850 hPa, hatches); streamlines of the 500 hPa and 850 hPa winds (purple); contours of temperature (black).

#### 2.1.4. CASIM Double- and Single-Moment Schemes With Height-Varying $RH_c$ (2M CS; 1M CS)

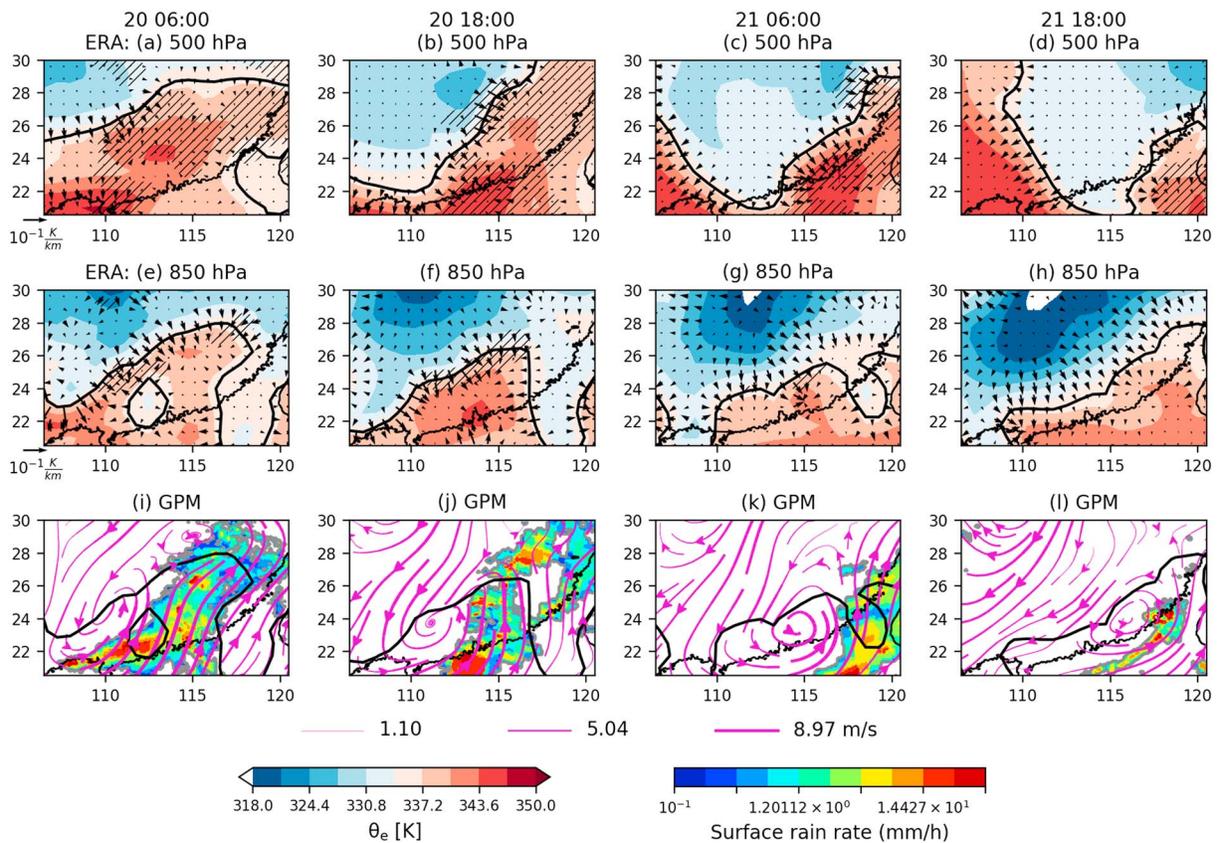
To assess the relative importance of the parameterization of condensation, compared to cloud microphysical processes, a pair of sensitivity tests to the parameterized values of  $RH_c$  are included. A vertical profile of values of  $RH_c$  is specified which transitions from  $RH_c = 0.95$ , near the surface, to 0.8 in the free-troposphere (above 5 km). In these simulations, cloud condensation occurs at lower values of relative humidity which means that less gridbox mean cooling/moistening is required to induce condensation of cloud. As noted above, the value of  $RH_c$  does not directly affect the ice cloud fractions in these experiments.

#### 2.2. Description of the Case Study

The case selected (hereafter, SR1) occurred between 19 and 21 May 2016 over Southern China. A squall line of organized deep convection formed inland of the south coast of China, and propagated eastward, over a 48-hr period, bringing heavy rain to southeastern coastal areas, before eventually traveling out over the South China Sea. The simulations of the case were initialized at 00 UTC on 20 May, allowing for 6 hr of *spin-up* prior to the first comparison with observations.

We will use a combination of the European Centre for Medium-Range Weather Forecasts’ ERA Interim reanalysis, composite-radar imagery and satellite-derived rainfall estimates to characterize the large-scale conditions that drove the squall line formation and propagation. Figure 1a shows the composite-radar reflectivity (maximum reflectivity in column) imagery, derived from the Chinese Meteorological Administration’s network of operational weather radars, for SR1 at 0800 BJT (00 UTC) on 20 May 2016. A band of heavy rain can be seen extending from Guangdong and Guangxi provinces, in the southwest, to Jiangxi and Fujian provinces in the east. The associated large-scale fields can be seen in Figures 1b and 1c, which show the dynamic- and thermodynamic-state of the middle (500 hPa, Figure 1b) and lower (850 hPa, Figure 1c) troposphere, in the ERA Interim reanalysis for 00 UTC on 20 May 2016. The streamlines of the horizontal wind components are shown, together with the equivalent potential temperature ( $\theta_e$ , color shading), and temperature contours (solid black lines). The hatched areas at 500 hPa and 850 hPa show the regions of negative pressure-velocity,  $\omega$  (Pa/s), and horizontal divergence,  $\nabla_h \cdot \mathbf{u}$  ( $s^{-1}$ ), respectively. These fields show that a strong southwesterly, low-level jet was present which brought warm, humid air from the South China Sea into contact with a cooler continental air mass, creating the conditions for synoptic and mesoscale forcing to generate a region of large-scale ascent and heavy rain formation. The region shown in Figures 1b and 1c corresponds to the integration domain used for the simulations, which was chosen to encompass the heavy rainfall over land.

The subsequent evolution of the squall line can be seen in Figure 2, which shows the reanalysis, Figures 2a–2h, and GPM IMERG surface rain rate (see section 2.6), Figures 2i–2l, at 6-hr intervals during the rainfall event. The reanalyses again show  $\theta_e$  (red-blue shading),  $\omega$  (at 500 hPa), and  $\nabla_h \cdot \mathbf{u}$  (at 850 hPa). Also shown



**Figure 2.** The evolution of the synoptic-scale conditions and surface rainfall associated with the SR1 event. (a–h) From the ERA Interim reanalysis:  $\theta_e$  (red-blue colors),  $\theta_e$  gradient vectors (black), regions of negative  $\omega$  (hatches, 500 hPa), and regions of negative  $\nabla_h \cdot \mathbf{u}$  (hatches, 850 hPa). (i–l) The GPM IMERG surface rainfall (color shading), streamlines of the 850-hPa wind from ERA Interim (purple) and  $\theta_e = 336$  K contours (black).

are the horizontal spatial-gradient vectors,  $\nabla_h \theta_e$ , of the on-pressure-level equivalent potential temperature, which indicate the regions of largest  $\theta_e$  gradients. These regions identify the location of the thermodynamic front in the lower atmosphere. The streamlines of the 850-hPa flow field are superimposed over GPM rain rates. For reference, a  $\theta_e = 336$ -K contour, which (for this case) coincides with the large-scale front on the 850 hPa, and 500 hPa, surfaces, is also shown (solid black lines). The thermodynamic front is seen to be associated with a mesoscale vortex at low levels (Figures 2i–2l) which moves from west to east across the domain, with the region of large-scale ascent and organized rainfall in the warm sector ahead of the vortex.

### 2.3. Description of the Data Sets Used for Model Evaluation

In section 3, the model sensitivity tests are evaluated against remotely sensed estimates of top-of-atmosphere radiative fluxes, cloud liquid water path, cloud top height, and surface rain rates. In addition, the radar reflectivity factors simulated by the model configurations are compared to measurements from a ground-based radar. The polar-orbiting satellite observatories used in this study are in Sun-synchronous orbits within the *A-train* satellite formation. They therefore pass over the case study region twice daily, once in the morning (around 2 a.m., local time) and once in the afternoon (around 2 p.m.). In this section, we will give an overview of the data sets used, discuss some uncertainty estimates obtained from previous studies, and describe the metrics used for model evaluation.

### 2.4. Clouds and the Earth's Radiant Energy System: Single-Scanner Footprint

The outgoing longwave (LW) and shortwave (SW) radiative fluxes at the top of the atmosphere are strongly influenced by clouds, aerosols, atmospheric thermodynamics, and surface properties. As such, these quantities provide a wealth of useful information for evaluating model processes. The Clouds and the Earth's Radiant Energy System (CERES) scanning-broadband radiometers, on NASA's Aqua and Terra satellites, provide retrievals of the broadband SW and LW fluxes at 20-km spatial resolution (Wielicki et al., 1996). In this study, we will use the Aqua edition 3A Single-Scanner Footprint (SSF) data. The CERES instrument measures

radiances in a SW-band, at 0.3–5  $\mu\text{m}$ , and in the infrared window at 8–12  $\mu\text{m}$ . The detected radiances are corrected to account for effects of the instrument's optics (Loeb et al., 2001), and angularly integrated in a way that empirically accounts for the scene-dependent anisotropy of the radiation field, to obtain estimates of the top-of-atmosphere radiative fluxes (Loeb et al., 2007). The resultant retrieval has estimated uncertainties of 5  $\text{W}/\text{m}^2$  in the SW and 2  $\text{W}/\text{m}^2$  in the LW (Loeb et al., 2007).

### 2.5. The Moderate Resolution Imaging Spectroradiometer

The Moderate Resolution Imaging Spectroradiometer (MODIS) is a spectrum-resolving scanning radiometer that operates between 0.4 and 14.5  $\mu\text{m}$  and has band-dependent resolutions ranging from 100 m to 1 km. The MODIS Cloud Product uses a combination of visible and thermal-infrared wavelengths, to retrieve estimates of the radiative and material properties of clouds. We will use the MODIS/Aqua Collection 6 Level-2 (derived geophysical variables at the *pixel*—that is, individual-observed-scene—level) estimates of cloud water path and cloud top pressure (Platnick et al., 2015). These fields are available in 5-min time interval *granules*, which individually cover an area of approximately  $1,354 \times 2,030 \text{ km}^2$ . Within a granule, the cloud water path product is available on a  $1 \times 1\text{-km}^2$  grid, and the cloud top pressure on a  $5 \times 5 \text{ km}^2$  that is geolocated at the central cell of each  $5 \times 5$ -block of 1-km cells. For midlevel and high-level cloud tops, the cloud top pressure is retrieved using a  $\text{CO}_2$ -slicing technique that exploits the variable opacity of the atmosphere in the  $\text{CO}_2$  absorption region (Menzel et al., 2008). Because the sensitivity of this method is highest in the upper troposphere, it is most effective for midlevel and high-level clouds. For multilayer clouds, it reports a pressure that is consistent with the radiative mean of the layers. The  $\text{CO}_2$ -slicing method also requires sufficient infrared-contrast between cloudy and clear skies and, hence, is less accurate for low cloud and optically thin high clouds. Consequently, for low clouds, the retrieval usually reverts to inferring cloud top pressure by combining an infrared-channel brightness temperature measurement with a model analysis.

The prelaunch evaluation by Frey et al. (1999) of an airborne simulator of MODIS against cloud lidar measurement indicates that biases of less than 1.5 km in cloud top height can be expected in more than two thirds of cases in which  $\text{CO}_2$  slicing is used. However, subsequent comparisons by Holz et al. (2008) to cloud top heights obtained from the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) instrument on board NASA's Cloud Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) platform showed that, although the two retrievals were relatively similar in terms of their detections of clouds, the estimated cloud top heights were sometimes highly discrepant. Of relevance to this study, Holz et al. (2008) reported that the top low-level clouds over the oceans differed by an average of 1 km, and those for high clouds differed by 4 km, in terms of a global mean, with extreme differences of 10 km in 16% of the nonpolar high-cloud cases, mainly as a result of the presence of optically thin clouds.

The presence of uncertainties in the estimation of low-cloud fraction are further supported by the land-based studies by Tang and Chen (2006) and Y. Li et al. (2006) who used the MODIS cloud properties (CTPs and optical thicknesses) to construct categorized climatologies of cloud types for the east Asian monsoon region (Tang & Chen, 2006) and over the Tibetan Plateau (Y. Li et al., 2006) and compared these to cloud types from a database of weather station reports. These studies found that MODIS underestimated the frequency of occurrence of most cloud types, with particularly large differences for low clouds; biases that they attributed to either the presence of thick high clouds obscuring the MODIS view, or low clouds obscuring high clouds from the ground-based observers.

For the SR1 case, the cloud field is composed of deep convective clouds, with low cloud top pressures, and low-to-mid level stratus and shallow cumulus clouds. The findings of Holz et al. (2008) suggest that the cloud top pressures retrieved for the optical thick deep clouds may be expected to fall in the relatively less uncertain part of the parameter space. However, the same may not be true for any optically thinner anvil cirrus that detrained from the periphery of the squall lines: such clouds might be misinterpreted as low-level or midlevel clouds by the MODIS retrieval.

The MODIS cloud water path is derived by combining estimates of cloud optical thickness and cloud particle effective radius (the ratio of second and third moments of the droplet size distribution), which are retrieved simultaneously from reflectance measurements in spectral bands that differ in terms of their amount of absorption by cloud particles. This procedure necessarily involves the use of a theoretical radiative transfer model (Platnick et al., 2017) and requires making pixel-by-pixel decisions as to which phase (liquid or ice) should be assumed. As such, MODIS retrieves a single condensed water path (CWP), that is, a column-integrated cloud water content, which may be indicative of liquid or ice depending on the cloud type.

As such, the retrievals are uncertain due to a combination of measurement and forward-model errors. For example, over the global oceans, studies have shown that the MODIS LWPs differ from microwave-based retrievals from the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E; Hilburn & Wentz, 2008). In particular, Greenwald (2009) and Seethala and Horváth (2010) showed that tropical and subtropical MODIS LWPs are, on average, negatively biased compared to AMSR-E, whereas the LWPs of high-latitude clouds are overestimated on average. Seethala and Horváth (2010) also identified a strong dependence of the biases on cloud regime, with the LWPs of warm stratocumulus and cumulus clouds being overestimated and underestimated, respectively. Moreover, comparisons to in situ measurements, for example, those by Painemal and Zuidema (2011) for the South Pacific marine stratocumulus decks, suggest that the source of the uncertainties is a systematic high bias in effective radius, which propagates into the LWP estimates. Comparisons over land show similar biases, for example, Hongru et al. (2015) compared CERES-MODIS LWPs to ground-based retrievals over the Loess (Huáng-tǔ) Plateau in central China and found mean biases of the order of  $40 \text{ g/m}^2$  for single layers of stratus clouds with liquid water paths in the approximate range  $50\text{--}200 \text{ m}^{-2}$ .

### 2.6. The Global Precipitation Measurement Surface Rainfall Retrievals

We will use surface rainfall estimates from the Global Precipitation Measurement (GPM) mission's Integrated Multisatellite Retrievals for GPM (IMERG) data set (Hou et al., 2014): a gridded-precipitation retrieval that calibrates information from the microwave sensors in the GPM virtual constellation of satellites, against the dual-frequency precipitation radar on the GPM Core Observatory, to provide 30-min,  $0.1^\circ \times 0.1^\circ$ , estimates of precipitation. The gridded product covers a latitudinal range from  $60^\circ\text{S}$  to  $60^\circ\text{N}$ . We will use the post real-time *final/research* version of IMERG, which is produced approximately 3 months after the observations are made and uses monthly rain gauge measurements to improve the precipitation estimates (Huffman et al., 2014).

Ning et al. (2016) evaluated the performance of IMERG against gridded daily rainfall analyses and quality controlled rain measurements over China. They found a mean bias of less than  $0.1 \text{ mm/day}$ , in daily mean rainfall rates during a 20-month comparison period. IMERG was also shown to be highly correlated with the observed rainfall, particularly outside of northern China and during the summer months. Of interest for the case studied here, they also showed that IMERG performed reasonably well (e.g., had a low relative bias and was highly correlated (Pearson's  $r$ -value above 0.9) with the observations) at capturing the frequency of occurrence of days with relatively heavy rainfall (when accumulations exceeded  $20 \text{ mm}$  in  $24 \text{ hr}$ ) and also the amount of rainfall on those days. Because it includes rain-gauge measurements, we expect IMERG to perform well (on average) over data-rich regions, but uncertainties for individual cases may nevertheless be large. The reliability of IMERG for individual cases of extreme rainfall during the east Asian summer monsoon was investigated by Wang et al. (2017), who evaluated IMERG for eight typhoons over the coastal China. They found relative biases of the order of  $10\%$  for rain rates in the range  $20\text{--}60 \text{ mm/day}$ ; above/below this range, IMERG underestimated/overestimated rainfall rates by approximately  $\pm 20\%$ .

### 2.7. Radar Measurements and Model Evaluation Metrics

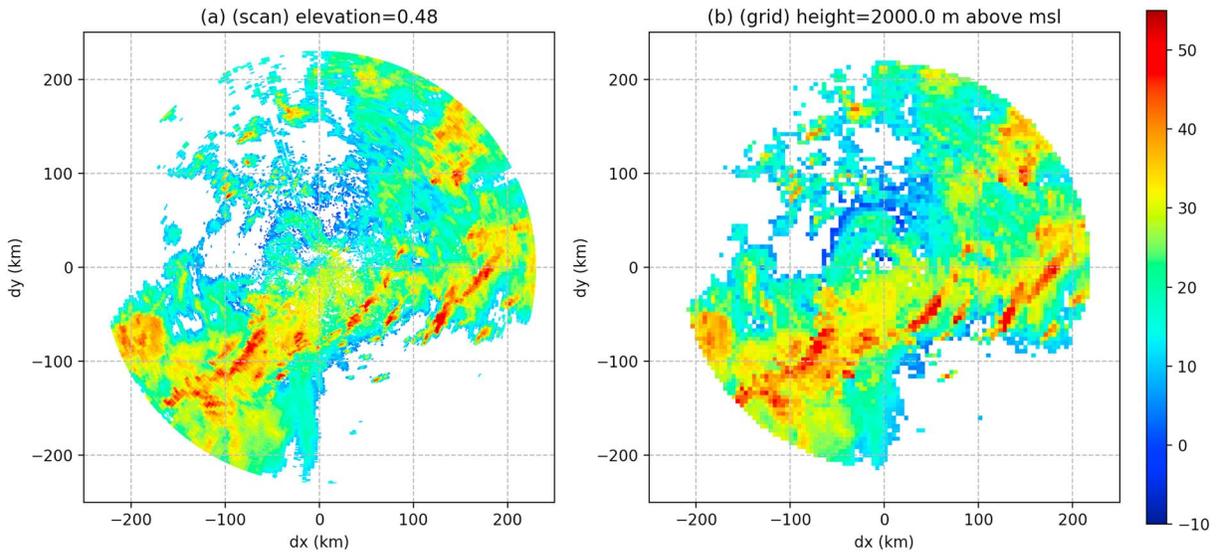
Measurements of radar reflectivity factor,  $Z$ , are obtained from an operational S-band radar located in Guangzhou at ( $113.35^\circ\text{E}, 23.00^\circ \text{N}$ ). The maximum range of the radar is approximately  $200 \text{ km}$  and volume-scans containing 9 elevations are available at intervals of 6 minutes. The azimuthal-resolution of the scans is 1 degree and there are 900 equispaced radial-gates. The data is available via the SCMREX website Luo et al. (2018).

Although beam can lead to nonlocal effects, a useful property of reflectivity at S-band wavelengths is that it depends strongly on the local Rayleigh-scattering properties of the population of hydrometeors, at the location of the radar echo, via the formula:

$$Z \propto \sum_{s \in S} \int dM_s^2 f_s, \quad (1)$$

where  $S$  is the set of types of hydrometeor, for example,  $\{\text{cloud}, \text{rain}, \dots\}$ ,  $f_s(D)$  is the particle size distribution for species  $s$ , and  $M_s$  is the particle mass.

This makes forward modeling radar reflectivity from the model output particularly simple because complex scattering cross-section calculations are not involved and the geometry of the model grid, relative to the radar, does not need to be accounted for. To compare the simulations to the Guangzhou radar, the observed reflectivity is interpolated, from the volume scan coordinate system of the radar, onto the model latitude-longitude



**Figure 3.** (a) The measured reflectivity factors on the lowest elevation scan of the radar at 09 UTC 20 May, compared to (b) the result of interpolating the reflectivity measurements onto the model latitude-longitude grid, at a fixed height of 2 km above mean sea level.

grid at fixed heights of 1, 2, and 5 km above mean sea level (MSL). Figure 3 shows the result of the regridding at a height of 2 km, compared to the lowest-elevation ( $0.48^\circ$ ) radar scan. The regridding is performed by three-dimensional linear interpolation in  $Z$  using the tetrahedral simplices obtained from a Delaunay tessellation within the convex hull of the radar observations. Hence, we use an arithmetic volume mean to represent the observed reflectivity in each model gridbox. For a discussion of the accuracy of this assumption (compared to using the geometric mean of the observations), see Lakshmanan (2012). We have not investigated the choice of interpolation method, because the differences between observations and models are large, and because artifacts due to regridding are expected to be smaller than unquantifiable uncertainties due to the spatial resolution of the radar and attenuation.

To compare the satellite observations to the model predictions, the measurements and predictions are both interpolated onto a latitude-longitude grid with a grid spacing specified by whichever data set has the lowest resolution. Hence, we compare radar, models, and MODIS on the model grid; CERES is compared to the models on a grid with a fixed spacing of  $0.2^\circ$  (approximately 22 km, that is, the estimated resolution of the CERES instrument at nadir). Where more than one MODIS granule, or CERES orbit, is required to fully cover of the case study domain, the required data points are composited prior to the regridding. This creates a slight ambiguity as to what constitutes the *best* validation time for comparing model to observations. To avoid complicating the analysis, we chose to evaluate the simulations at the hour that occurs most frequently in each composite satellite scene. In practice, this means that a maximum time difference of 1 hr may exist between model and observations for some pixels within the domain. Because the CERES and MODIS observations are collocated on the model grid, it is possible to partially decompose the model radiative flux biases according to cloud type. In this study we will use cloud top pressure to examine how LW and SW flux biases are distributed between high, medium, and low clouds, according to the International Satellite Cloud Climatology Project (ISCCP) classification. Given a top-of-atmosphere radiative flux,  $F$ , from a given data source (from either a model or observations, CERES), and the corresponding cloud top pressure field (model or MODIS), the domain mean flux,  $\langle F \rangle$ , can be decomposed into contributions due to grid points for which the cloud top pressure,  $p_c$ , falls within each of a specified set of ranges. For the ISCCP classification, these ranges are as follows:  $p_c > 680$ ,  $680 > p_c > 440$ , and  $p_c < 440$ , for low, medium, and high cloud, respectively. Hence,

$$\langle F \rangle = \sum_i A_i \langle F \rangle_i, \quad (2)$$

where  $A_i$  is the fraction of the area of the domain for which  $p_c$  satisfies the conditions for cloud category  $i \in \{\text{low, medium, high}\}$  and  $\langle F \rangle_i$  is the area-averaged flux for that category.

To analyze the effects of the microphysics configurations on the simulated clouds, it is convenient to separate changes in in-cloud quantities from changes in cloud cover. To do this, a subvolume of the computational

domain needs to be defined as *cloudy*. In this study, we use the expressions (bulk-)cloud fraction and frequency, interchangeably, to refer to the average frequency of occurrence of cloud, at a given height and time, defined as the fractional area of the computational domain in which the condensed water content exceeds a specified threshold,  $q_{s0}$ , (taken to be  $10^{-8}$  kg/kg):

$$\phi_s = \sum'_{\mathbf{x}|z=h} C_s(\mathbf{x}) \frac{\Delta(\mathbf{x})}{A}, \quad (3)$$

where  $C_s$  is the subgrid-scale cloud fraction of species  $s$ ,  $\Delta(\mathbf{x})$  is the area of the grid box centered at  $\mathbf{x}$ ,  $h$  is the value of the vertical coordinate ( $z$ ) at which the average is taken, and  $A$  is the total area of the domain. The prime (') symbol on the summation sign ( $\sum'$ ) indicates that the sum is taken only over points for which  $q_s(\mathbf{x}) > q_{s0}$ .

Similarly, by the *in-cloud water content*, or *concentration*, of a hydrometeor species, we mean the average of the condensed water content,  $r_s = \rho_a q_s$  ( $\rho_a$  is the air density), over the subvolume of the domain where  $q_s$  exceeds the threshold for defining the presence of cloud, that is, the total above threshold condensed water content, divided by the area covered by cloud:

$$\rho_s = \sum'_{\mathbf{x}|z=h} r_s \frac{\Delta(\mathbf{x})}{A\phi_s}, \quad (4)$$

Note that this definition differs slightly from the area average of the in-cloud condensed water content; the latter quantity would be  $\sum_{\mathbf{x}}(r_s/C_s)\Delta/A$ , and will not be used in this study. Finally, the *mean water content* or (cloud-)amount refers to

$$\alpha_s = \sum'_{\mathbf{x}|z=h} r_s \frac{\Delta(\mathbf{x})}{A}. \quad (5)$$

Hence,  $\alpha_s = \phi_s \rho_s$  and therefore contains contributions from both the area covered by clouds and the condensed water contained within those clouds. We will also make use of condensed water paths, which are column-integrated cloud amounts and are defined as  $W_s = \sum_z \alpha_s(z) \Delta_z$ , where  $\Delta_z$  is the vertical depth of a gridbox.

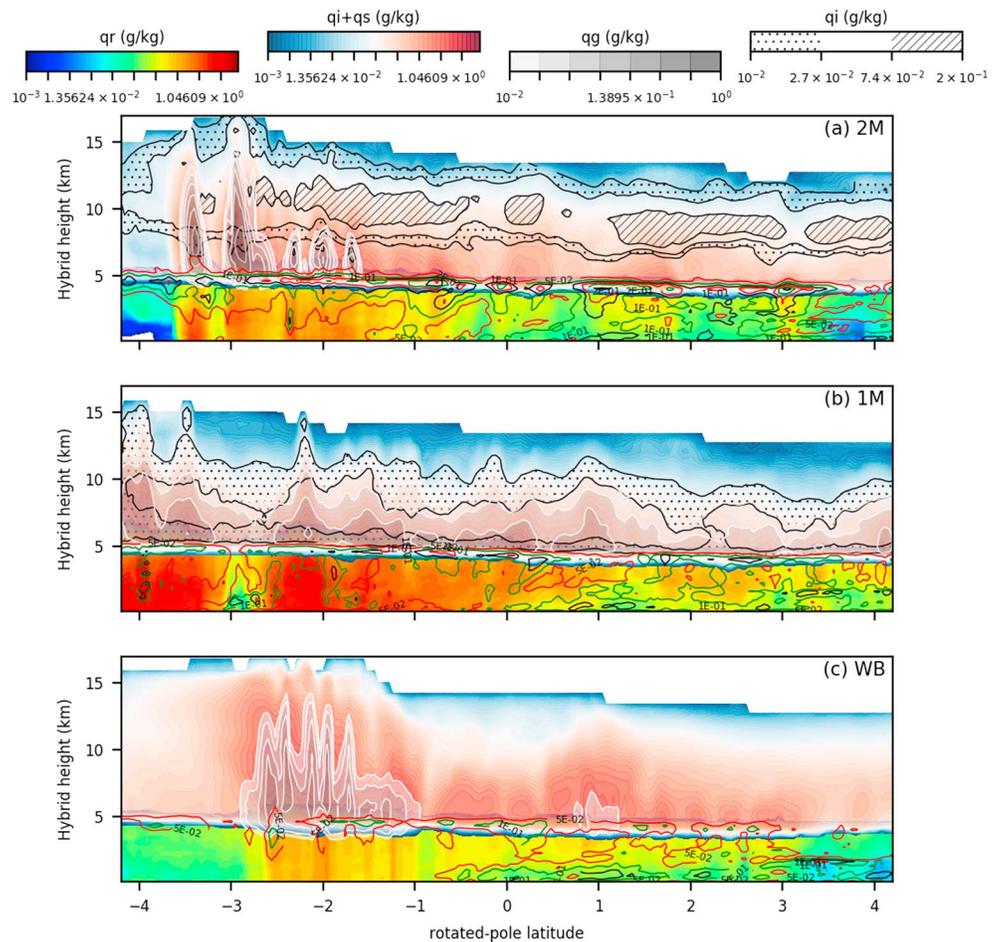
When we compare the models to the observations of radar reflectivity, we will find that a pointwise evaluation of model errors is not informative because forecast errors, within the relatively small sampling volume of the radar, mean that the simulated and observed clouds are usually not collocated. Hence, a statistical comparison of predicted and observed reflectivity is more informative for understanding model sensitivity. To this end, we will employ three statistical metrics to describe the frequency distribution of radar reflectivity which, although deliberately unsophisticated, are sufficient to identify the main biases and intermodel variations. In addition, we must account for the physical limitations of the radar when comparing low-intensity echos and also for the irregular shape of the radar-scan volume when defining a subdomain for the comparison. Because the radar is known to have very low sensitivity to returns below 5 dBZ, we will use this value as a lower bound when analyzing the model data. For the comparison volume, we will use the convex hull of points defined by the interpolation procedure. Within this volume, the statistical metrics used are as follows: the median reflectivity, for reflectivity values above the 5-dBZ threshold; the *skew*,  $s_k$ , of the frequency distribution,  $h(Z)$ , which we define as

$$s_k = h(30 \text{ dBZ}) - h(10 \text{ dBZ}), \quad (6)$$

that is, the difference in the frequency of gearboxes between 30 dBZ and 10 dBZ; and the *overlap* of a distribution with the observations,

$$O_h = \sum_{Z > 5 \text{ dBZ}} h(Z)h_{\text{obs}}(Z) / \sum_{Z > 5 \text{ dBZ}} h_{\text{obs}}^2, \quad (7)$$

where  $h_{\text{obs}}$  is the frequency distribution of the radar observations. The last of these metrics describes (roughly speaking) how well the simulated values *project* onto the observations, in dBZ space. In particular,  $O_h$  measures how often a simulation produces the reflectivities that occur with the highest frequencies in the observations. (A model that was statistically perfect, in this regard, would have  $O_h = 1$ , whereas  $O_h > 1$  and  $O_h < 1$  indicate models that overestimate and underestimate, respectively, the frequencies of the most frequently occurring observed values.)



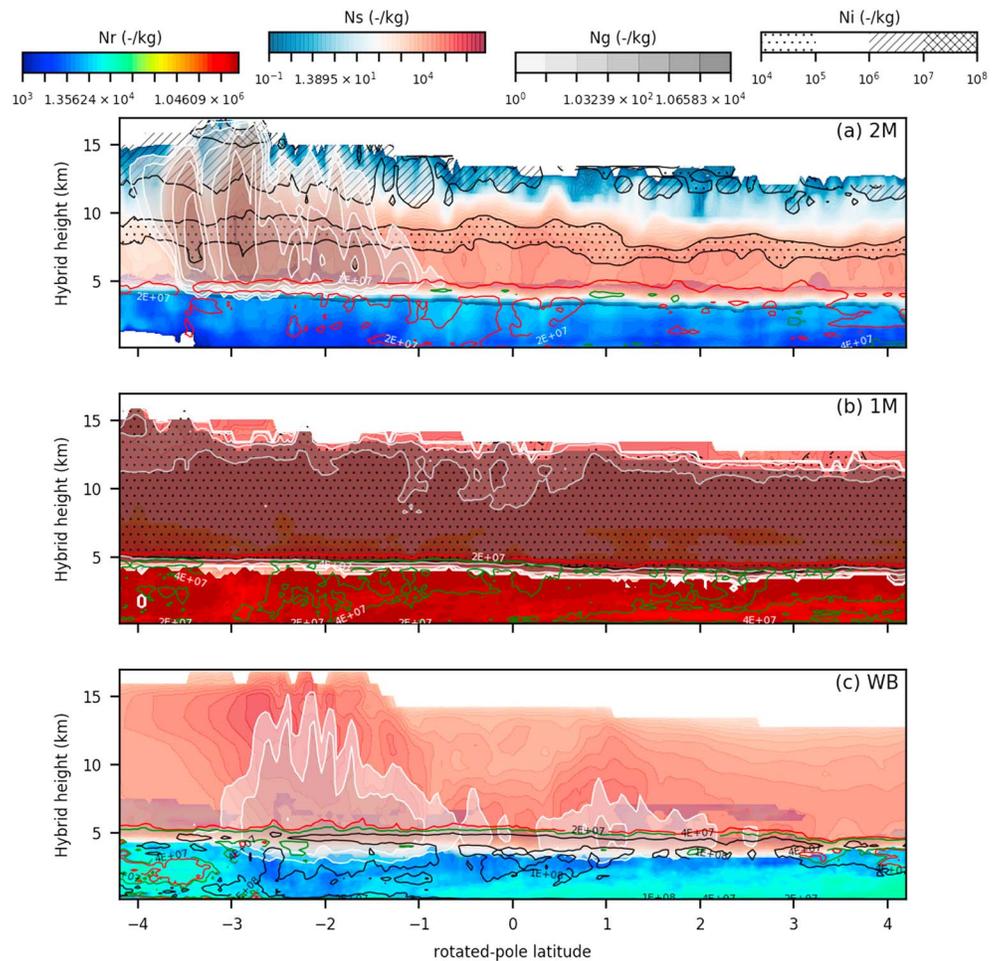
**Figure 4.** Zonally averaged vertical cross sections of condensed water content for the (a) 2M, (b) 1M, and (c) WB configurations. The colors show rain water content, total ice water content, and graupel water content, according to the color scales shown at the top of the figure. The contour lines are lines constant cloud water content (colored) and cloud-ice water content (black). The hatches show the cloud-ice mass mixing ratio, according to the scale at the top of the figure. The fields are plotted in the intrinsic coordinate system of the model: the vertical axis is the model hybrid height coordinate (essentially, this is equal to height above surface in the lower atmosphere and transitions to height above mean sea level aloft); the horizontal axis is latitude in the rotated-pole coordinates.

### 3. Case Study Results

We consider the WB, double- and single-moment CASIM configurations to be *baseline* model configurations, the differences between which we aim to understand. In sections 3.1 and 3.2, as a first step to quantifying these differences, we consider the bulk (average) properties of the cloud and precipitation fields in these simulations. These properties include the mean vertical structures of the simulated cloud fields (section 3.1) and the spatiotemporal characteristics of the surface rainfall rates (section 3.2). At the same time as identifying the main interconfiguration differences in model predictions, we will discuss results from the *intermediate* model configurations listed in Table 1. These configurations are shown to reproduce most of the variations in cloud structure and some of precipitation variability. Subsequently, the effects of the model configurations on errors in clouds and radiation are evaluated in sections 3.3 and 3.4. Finally, in section 3.5 we summarize the main factors among the scheme components tested that influence the biases.

#### 3.1. Average Cloud Structures: Cross Sections and Vertical Profiles

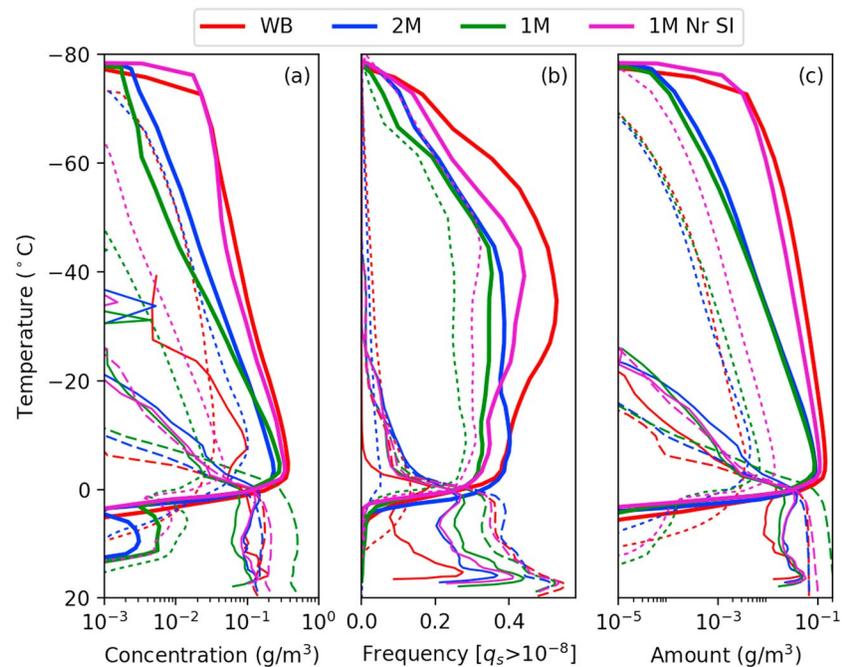
The main differences between the simulated cloud fields in the baseline configurations can be assessed from zonally averaged vertical cross sections through the SR1 squall line. Examples of these cross sections are shown in Figure 4 at 00 UTC on 21 May 2016. The three simulations produce noticeably different predictions of condensed water content: both the amounts and spatial distributions of the hydrometeor fields vary with model configuration. For example, the WB scheme produces more ice water than the CASIM schemes,



**Figure 5.** Zonally averaged vertical cross sections of hydrometeor number density for the (a) 2M, (b) 1M, and (c) WB configurations. The colors show the number concentrations of raindrops, snow aggregates, and graupel particles, according to the color scales shown at the top of the figure. The colored contour lines are lines constant cloud droplet number concentration. The hatches show the cloud ice number concentrations, according to the scale at the top of the figure. The fields are plotted in the intrinsic model coordinate system of the model: the vertical axis is the model hybrid height coordinate (essentially, this is equal to height above surface in the lower atmosphere, and transitions to height above mean sea level aloft); the horizontal axis is latitude in the rotated-pole coordinates.

but gives rain water contents (RWCs) that appear qualitatively comparable to the 2M scheme, whereas the rain mass from the 1M scheme is larger than in either of the other two simulations. There are also differences in the graupel generated by the simulations: in 2M and WB the graupel is concentrated within the regions of highest cloud water content; by contrast, the 1M scheme produces a more horizontally uniform distribution of graupel.

The differences in condensed water contents are related to the differences in the number concentrations of hydrometeors. Figure 5 shows the number concentrations of each hydrometeor species that are prognosed (in the case of 2M) or diagnosed (in the case of single-moment schemes) in the simulations. Several obvious differences stand out: first, the number concentrations of all species are higher in 1M than in 2M or WB; second, the trend of rain number concentration with height in WB is the opposite of that seen in CASIM simulations. The number concentrations in 1M are empirically determined constants, so it is interesting that the values specified are much larger than those that evolve prognostically in the double-moment scheme. Below, it will be shown that this difference is a substantial contributing factor to the rain water contents and radar reflectivity factors for rain, in the models. Similarly, the WB scheme uses a variable-intercept parameterization that apparently responds very differently to the rapidly reducing mass of rain in the subcloud layer, compared to the other two schemes (Abel & Boutle, 2012). The behavior of the DSD parameterization in WB has been

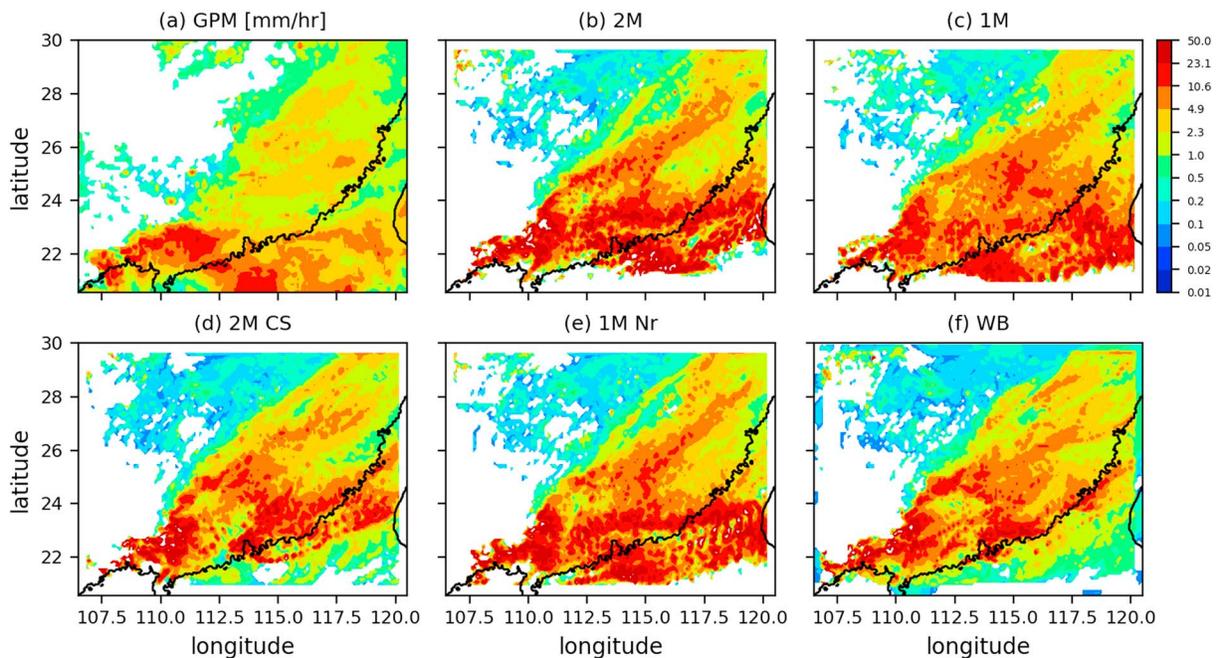


**Figure 6.** The area-averaged vertical profiles of (a) in-cloud condensed water (concentration),  $\rho_s$ , defined as the mean mass of condensate per unit volume averaged over all cloudy grid boxes; (b) bulk cloud fraction (frequency),  $\phi_s$ , defined as the fraction of the domain area that is occupied by cloud at each height; (c) condensed water amount,  $\alpha_s$ , defined as the average mass of condensate. The colors denote different model configurations, according to the legend at the top of the figure. The various styles of line denote different species of condensate: cloud liquid (thin-solid lines); rain (dashed lines); graupel (dotted lines); cloud ice and snow (thick-solid lines).

chosen to transition from frontal environments to drizzling stratocumulus, as the RWC decreases. Figure 5c suggests that the DSD also undergoes this transition in response to evaporation of rain, which, in this case, produces results that are at odds with the 2M simulations.

We can quantify the differences in the hydrometeor fields by using the vertical profiles of in-cloud water content (concentration), cloud fraction (frequency of occurrence), and cloud amount, shown in Figure 6. The lines show the domain mean profiles, for each hydrometeor species, as functions of temperature. Below the melting level ( $T > 0^\circ\text{C}$ ) the main interconfiguration difference is the larger in-cloud amount of rain water in 1M (green-dashed line; Figure 6a), which leads to a correspondingly larger rain amount (Figure 6c), compared to the other simulations. Above the melting level, the biggest difference is that the WB simulation (solid-red lines) has a larger amount of ice cloud, both in terms of cloud fractions and in-cloud ice water contents (IWCs). By contrast, the single- and double-moment CASIM simulations are relatively similar in terms of the prevalence and concentration of the ice clouds. Note that for the CASIM schemes, the thick solid lines show the *total* ice cloud, including the crystal and aggregate species. In addition to these main differences, there are noticeable *second-order* variations in the graupel water contents and the structure of the mixed-phase layer. In particular, the graupel structures in the 2M and WB simulations exhibit higher densities of graupel, concentrated into smaller areas, compared with the corresponding structures in 1M. Liquid water also extends to a greater height in the WB simulations.

The differences in ice clouds between the CASIM 1M and WB simulations are investigated by using the 1M Nr SI sensitivity test. The later configuration uses single-moment CASIM, but with modifications to the ice microphysics so that the scheme more closely resembles WB. In particular, 1M Nr SI uses a single hydrometeor category for both snow and ice, uses the same mass-diameter relationship as the WB scheme, and uses the same parameterization of mass-weighted mean fallspeed as the WB scheme. It can be seen that this combination of factors accounts for most of the differences in ice clouds between 1M and WB. Moreover, further sensitivities (not shown) demonstrate that among these changes the single most important factor, for this case, is the modified sedimentation flux. In particular, the use of one prognostic ice species has only a small



**Figure 7.** Time-averaged spatial distributions of surface rainfall rate for the GPM IMERG satellite product (a) and the 2M (b), 1M (c), 2M CS (d), 1M Nr (e), and WB (f) model configurations. The colors show the mean rain rates, in mm/hr; the black contour shows the coastline of south China.

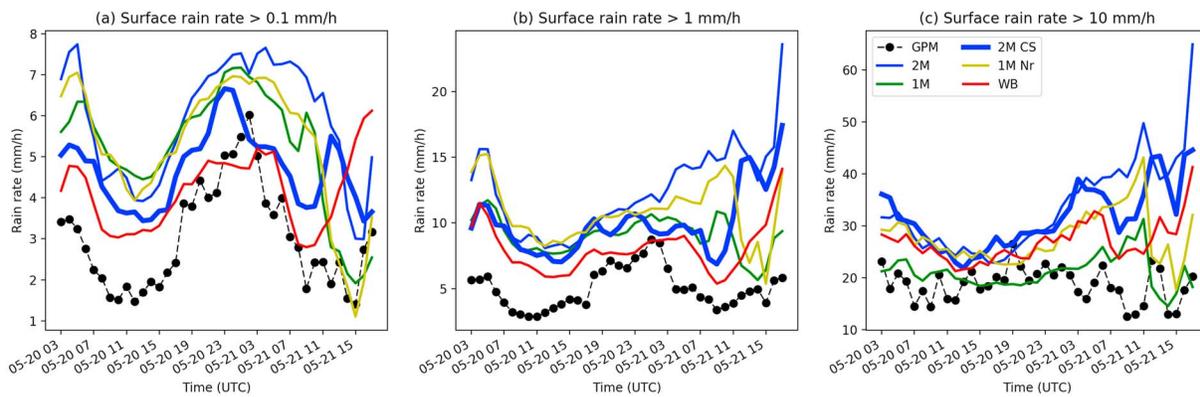
affect, since the amount of cloud ice in the CASIM simulations is small compared to the amount of snow, and the microphysical properties of the two species are similar.

The 1M Nr SI experiment can also reproduce the lower rain water contents seen in CASIM 2M and WB. These are strongly influenced by the diagnosed number concentration of rain in 1M, which is considerably larger than the number concentrations in either WB or 2M. Reducing the number concentration in 1M to resemble the values predicted by the other two schemes reduces the rain water contents so that they become similar to 2M and WB.

### 3.2. Evolution and Statistics of Surface Rainfall

There is some evidence from Figure 4 that the choice of cloud microphysics scheme influences the macrodynamical evolution of the squall line. In particular, in the 2M simulation (Figure 4a), there is a region of intense convective activity, between 2° and 4° away from the center of the domain, that is displaced by several hundred kilometers from its location in the WB simulation (Figure 4a). This suggests that the cumulative effects of differences in diabatic-heating rates, between the simulations, are sufficient to induce variations in the propagation of convective features within the storm. To further examine these variations, Figure 7 shows the mean precipitation structures associated with a subset of the sensitivity tests and compare these to those seen in GPM IMERG. The time-averaged rain rates, shown in Figure 7, demonstrate that the observed rainfall patterns, as estimated by the satellite product (Figure 7a), are qualitatively similar to those generated by the CASIM simulations (Figures 7b–7e). The WB simulations resemble the satellite retrievals over the land, but apparently underestimate the amount of rainfall over the South China Sea. There are also noticeable, but smaller, differences in simulated rainfall between the CASIM 2M (Figure 7b) and 1M (Figure 7c) simulations, which will be evaluated quantitatively below. Moreover, Figure 7e (1M Nr) suggests that these differences can, to some extent, be attributed to the differences in raindrop number concentrations,  $N_r$ , between the simulations. Figure 7d (2M CS) shows that  $RH_c$  also affects the surface rain rates in the double-moment simulation, resulting in a reduction in rainfall particularly over ocean.

Figure 7 suggests that, although there are differences in the finer-scale features of the rainfall, the large-scale evolution of the rain band is less affected by the choice of microphysics. This is consistent with the synoptic situation described in section 2.2, for which there is a substantial degree of control by large-scale thermodynamic gradients (see Figure 2). It is difficult, in this case, to attribute the evolution of smaller-scale features within the rain band to specific dynamical causes. It is however possible to demonstrate the potential for the evaporation of rain to affect the simulations. In particular, an analysis of the time-series of the rate



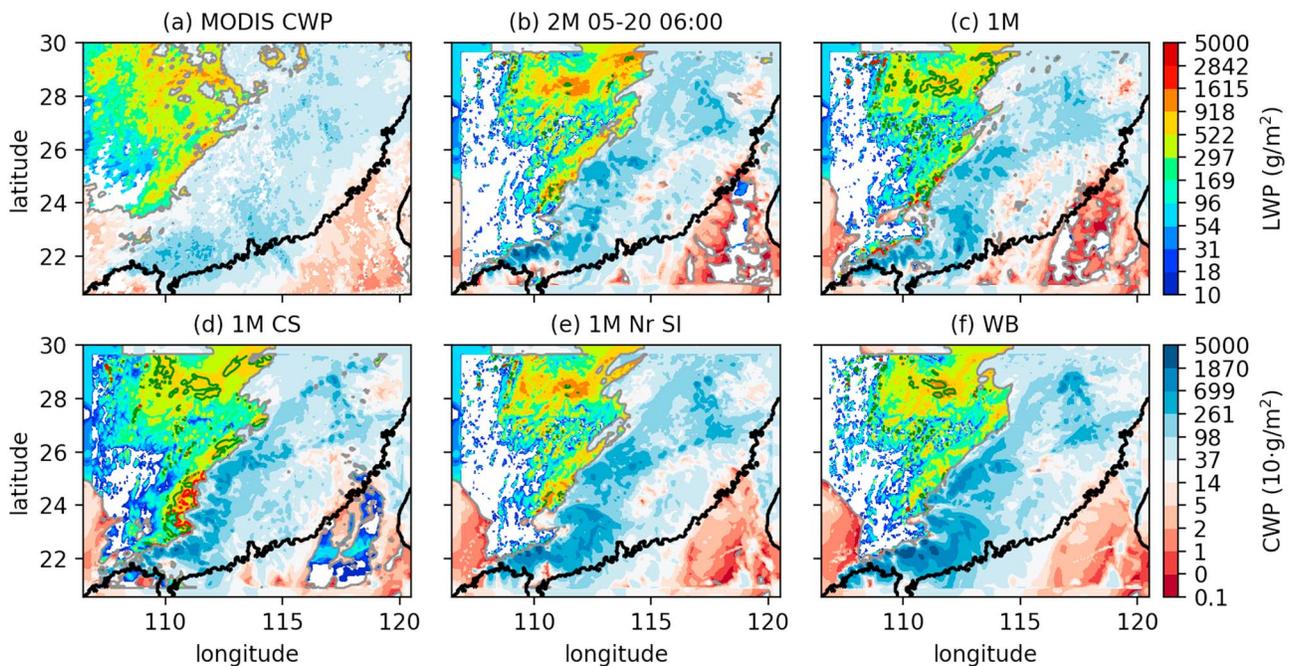
**Figure 8.** Time series of the domain-averaged surface rainfall rates, in mm/hr, for the duration of the case study. The rainfall rates over thresholds of (a) 0.1 mm/hr, (b) 1 mm/hr, and (c) 10 mm/hr, are shown for five model configurations and the GPM IMERG satellite product.

of evaporation of rain in the lower troposphere (figure not shown) shows that evaporation proceeds a factor of 2.5 faster in the 1M simulation, compared to 2M. This is consistent with the larger rain water contents produced by that configuration (see Figure 6). We have also identified a corresponding decrease in the average temperature of the subcloud air underneath the squall line, which (although it provides only a crude measure of cold-pool strength) suggests a potential feedback of rainfall on the properties of cold pools generated by convection within the squall line. In this case, the average temperature difference is very small (less than 1 K), which is consistent with the effects of evaporative cooling on the large-scale evolution being minor. However, the presence of a large-scale frontal boundary and the complexity of the underlying terrain may make it difficult to diagnose any sensitivity of cold pools to microphysics by considering average temperatures.

The qualitative differences in rainfall, seen in Figure 7, can be quantified in terms of their effects on the basic statistical properties of surface rain rate. To this end, Figure 8 compares the time series of simulated domain mean rainfall amounts to the GPM (and TRMM) retrievals, for rain rate thresholds of 0.1, 1, and 10 mm hr<sup>-1</sup>. For the lowest two thresholds, WB (red lines) gives the least surface rainfall and 2M the most. The single-moment CASIM scheme is intermediate between these two limits, except for the highest threshold, above which 1M has the least rain. Figure 8 also supports the conclusion that some of the rainfall variability, particularly above the higher two thresholds (Figures 8(b,c)) is due to the parameterization of  $N_r$ . Modifying this parameter in the single-moment CASIM scheme, gives rain rates (yellow lines) that are closer to the 2M scheme. The sensitivity of rain rate to cloud fraction diagnosis is also shown: the thick blue lines show that the rain rate decreases when  $RH_c$  is reduced. In terms of evaluation of the simulations against the satellite retrievals, the configurations consistently overestimate the average rain amounts, compared to GPM at all three rain rate thresholds.

### 3.3. Liquid Cloud Properties and the Radiative Impacts of Clouds

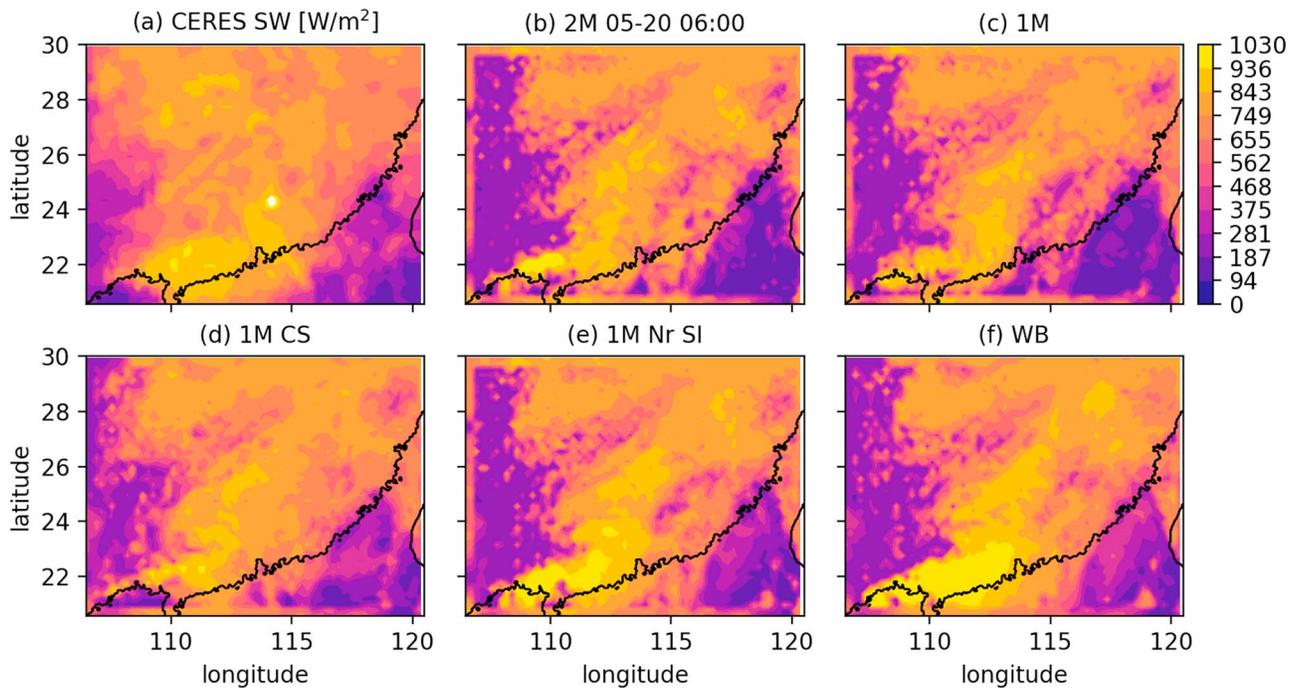
We can also examine the sensitivity of the spatial distribution of cloud liquid to the choice of microphysics scheme. Figure 9 compares the cloud liquid water path in the simulations to the MODIS retrieval at 06 UTC on 20 May. At this time, the squall line corresponds to the region of high water paths extending from southwest to northeast, across the domain. The MODIS retrievals treat cloudy pixels differently according to whether they are diagnosed as containing liquid, ice, or mixed-phase clouds. Hence, we partition the domain into two regions: a *warm*-phase region, containing only liquid-phase condensate; a *cold*-phase region, containing frozen-water mass. The MODIS retrieval can be partitioned using the cloud-phase flag generated by the retrieval algorithm (Platnick et al., 2015). For the simulations we define the cold-phase region as where the water path of frozen condensate (i.e., of cloud ice, snow, and graupel) is greater than 1 g/m<sup>2</sup>. The boundaries between the warm and cold regions are shown by gray contours. For the models, we have plotted the simulated cloud-liquid water paths in the warm region and the total condensed water path (including all hydrometeor types) in the cold region. For visual clarity, we use different color scales for the condensed water paths in each region—although, for the MODIS retrieval the same underlying field (the MODIS cloud water path) is plotted in both regions. For the models, it is convenient to indicate the areas of the warm region where rain and cloud droplets coexist. This is done by the colored contour lines in Figures 9b–9f). It can be seen that clouds to the northwest of the squall line are mainly warm-phase stratiform and small cumulus clouds with relatively little rain water.



**Figure 9.** The spatial distributions of cloud liquid water path, over the case study domain, at 06 UTC on 20 May 2016, from the MODIS satellite retrieval (a) and the 2M (b), 1M (c), 1M CS (d), 1M Nr SI (e), and WB (f) configurations. For the models, the colors show cloud liquid water path in  $\text{g}/\text{m}^2$  (according to the color scale shown top right), in the region where the frozen water path is less than  $1 \text{ g}/\text{m}^2$ ; total condensed water path (all phases; lower-right color scale), in the rest of the domain. For MODIS, the retrieved total condensed water path (CWP) parameter is plotted with a color scale chosen according to whether the retrieval determines the cloud as being liquid-only, or as containing ice. The gray contour lines designate the boundaries of the ice-containing regions. The blue, green, and red contours show (for the models only) the portions of the ice-free clouds where the simulated rain water paths exceed 10, 100, and  $1,000 \text{ g}/\text{m}^2$ , respectively. The black contours show the coastline of south China.

For the SR1 case the amount of liquid cloud is relatively insensitive to microphysics, although some small dependencies are noticeable. For example, the 2M and 1M CASIM simulations (Figures 9b and 9c) have more liquid cloud than the WB simulations, and double-moment microphysics produces slightly denser liquid clouds than the single-moment schemes. Modifying the raindrop number density,  $N_r$ , in the 1M scheme (Figure 9e) gives liquid clouds, which more closely resemble the 2M results, suggesting the differences between 1M and 2M are a result of differences in rain cloud accumulation rates. All these differences are, however, small compared to the effects of the cloud-fraction diagnosis (Figure 9b): using  $RH_c < 1$  in the 1M simulations, significantly enhances the area covered by liquid cloud. Because MODIS is sensitive to both liquid and ice the retrieved CWP is least ambiguous for the warm-phase clouds in the northwest quadrant of the domain, behind the squall line. Moreover, uncertainties in the retrievals mean that the model results should be compared cautiously to the MODIS retrieval. Nevertheless, it appears that by increasing liquid-cloudiness in the ice-free part of the domain, 1M CS gives a result that is more similar to the MODIS retrieval (although still comparatively lacking in cloud liquid). The spatial extent of the region containing cloud ice is similar in the simulations and MODIS, but the models all overestimate the highest condensed waters within the squall line compared to the MODIS estimates.

The differences in cloud species amounts, between the simulations, can be evaluated in terms of their effects on top-of-atmosphere radiative fluxes. Changes in the amount and composition of clouds are expected to influence the shortwave radiation, reflected back to space, and modulate the upwelling flux of terrestrial radiation. For example, the large increases in liquid cloud amount seen in Figure 9d are due to changes in cloud fraction diagnosis and therefore suggest that the outgoing SW flux, in particular, should be highly sensitive to the choice of  $RH_c$ . By contrast, ice cloud amount is not directly affected by  $RH_c$  (in the experiments considered here), but Figure 6 shows that the differences in ice microphysics between the WB and CASIM simulations give rise to different amounts of ice cloud. These changes should be detectable in the outgoing fluxes of radiation as changes due to the presence of cold, highly reflective, high cloud tops. Figures 10 and 11 show the TOA outgoing SW and LW fluxes, respectively, in the simulations and the CERES observations (Figures 10a and 11a), at 06 UTC 20 May 2016. In the simulations with more ice cloud (WB and 1M Nr SI) a larger area of the domain



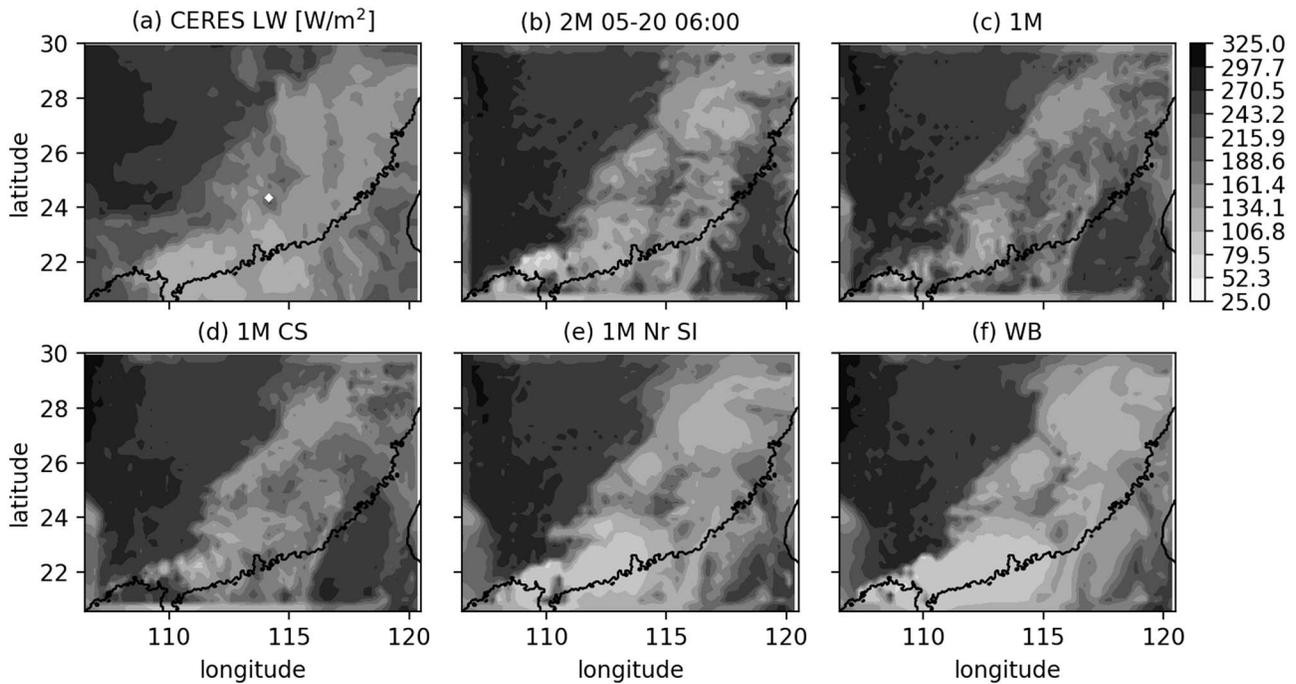
**Figure 10.** The spatial distributions of top-of-atmosphere, outgoing shortwave radiation (SW flux), over the case study domain, at 06 UTC on 20 May 2016, from the CERES broadband-flux measurements (a) and the 2M (b), 1M (c), 1M CS (d), 1M Nr SI (e), and WB (f) configurations. The colors show outgoing-SW flux in  $W/m^2$ ; the black contour shows the coastline of south China. CERES = Clouds and the Earth's Radiant Energy System; LW = longwave; SW = shortwave.

has high solar reflectivity (Figures 10e and 10f) and low thermal emissivity, compared to the simulations that have less ice cloud (2M, 1M, and 1M CS). The effect of  $RH_c$  can be seen by comparing Figures 10c and 10d, which shows the increase in SW flux due to the additional liquid cloud obscuring more of the underlying land surface. These effects can also be seen in the outgoing flux histograms shown in Figure 12. The SW flux histograms show two main sources of error, compared to the observations: an excess of relatively *dark* pixels, on the low-intensity side of the histograms; and biases in the frequency of relatively *bright* pixels, at higher intensities. The low-intensity bias is reduced by using  $RH_c < 1$  (which removes the spurious *peak* around  $200 W/m^2$ ), whereas the amount of ice cloud affects the errors at higher reflected intensities. Overall, it is difficult to identify schemes that perform substantially better than the others across both histograms. However, it will be shown in section 3.5 that the simulations with more ice cloud (WB and 1M Nr SI) and more liquid cloud (1M CS) have the smallest domain-averaged biases in TOA fluxes.

### 3.4. Sensitivity of Radar Reflectivity Factor to Microphysics

The amounts and number concentrations of the hydrometeor species also cause variations in radar reflectivity. In the Rayleigh-scattering regime, a radar echo, returned from a given volume of cloud, is determined by the  $M_5^2$  moments of the size distributions of the cloud particles. As such, reflectivity factor is strongly affected by particles of relatively large size or mass. Hence, condensed water content and number concentration assert opposing influences over the magnitude of the reflectivity. More mass of condensate, within a given volume, for a fixed number concentration, implies larger reflectivity. Conversely, for a fixed mass, a dense population of small particles will have a lower reflectivity.

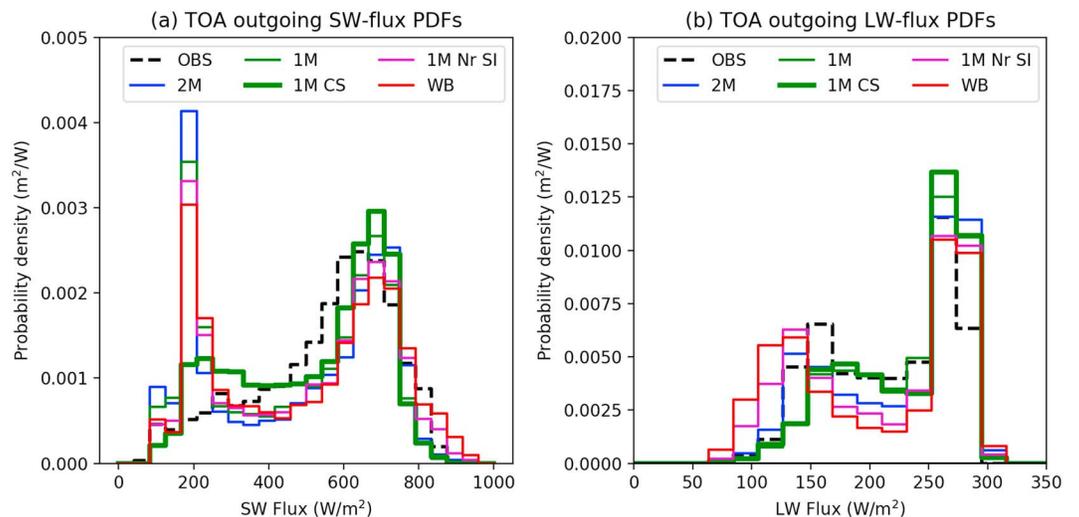
To evaluate these effects, Figure 13 shows the reflectivity factors, simulated by four of the model configurations, at heights of 2, 5, and 8 km above mean sea level and compares these to the measurements from the Guangzhou radar (Figures 13a–13c). Only a subregion, in the immediate vicinity of the radar is visualized. There are relatively large errors in the location of the simulated rainfall inside the volume of the radar scan. Moreover, these errors are consistent across the model configurations and can, therefore, reasonably be described as *cloud scheme-independent* forecast errors. Superimposed on these errors are the effects of the choice of microphysics scheme on the magnitude of the reflectivity. In particular, the 1M scheme stands out as having anomalously low reflectivity at low levels (2 km, Figure 13g), compared to the other schemes. These differences are consistent with the observation that the number concentration of rain in 1M is much larger



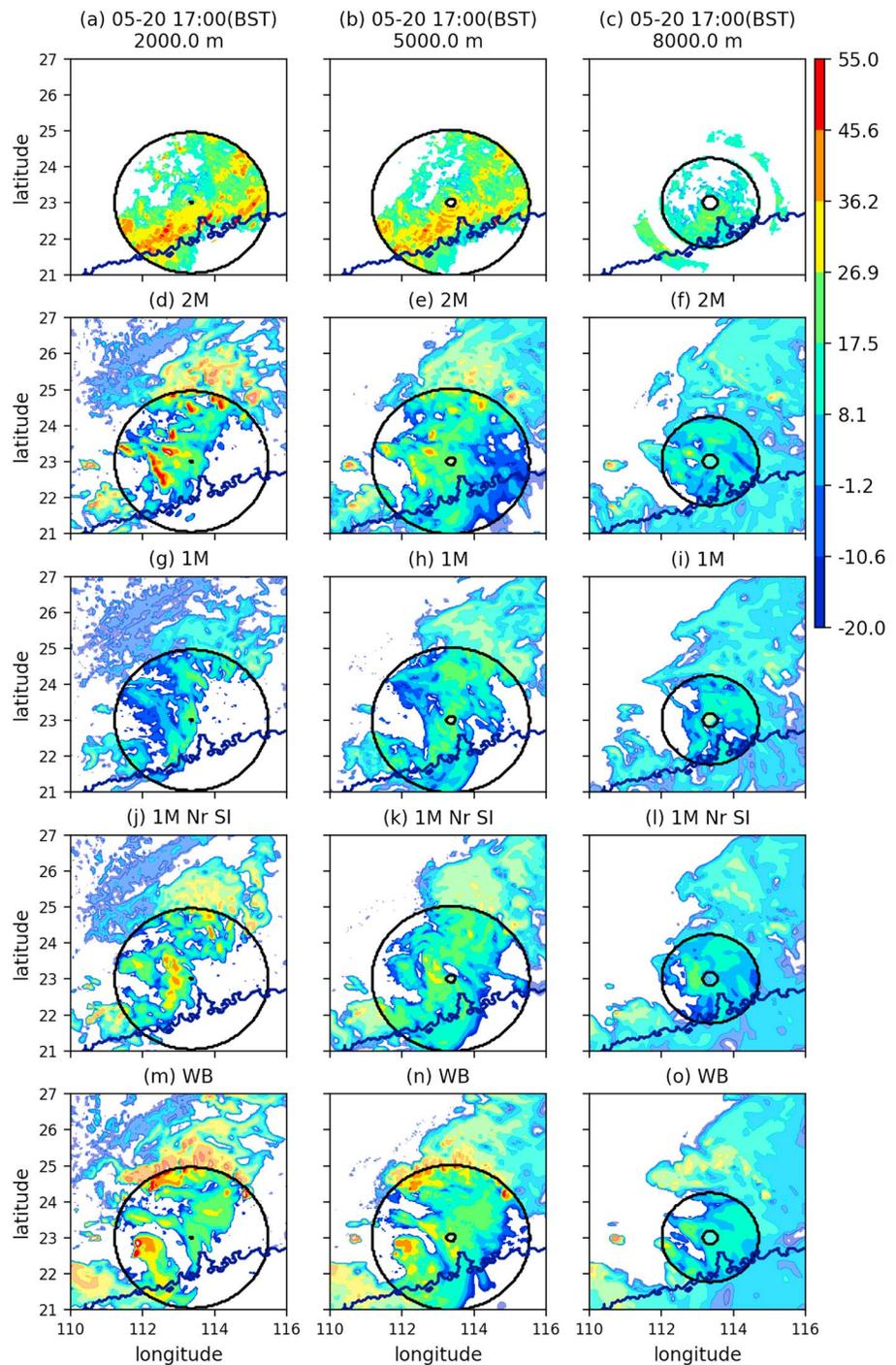
**Figure 11.** The spatial distributions of top-of-atmosphere, outgoing longwave radiation (LW flux), over the case study domain, at 06 UTC on 20 May 2016, from the CERES broadband flux measurements (a) and the 2M (b), 1M (c), 1M CS (d), 1M Nr SI (e), and WB (f) configurations. The colors show outgoing-LW flux in  $W/m^2$ ; the black contour shows the coastline of south China. LW = longwave; CERES = Clouds and the Earth’s Radiant Energy System.

than the corresponding value prognosed by the 2M scheme. Moreover, the 1M Nr SI simulation gives similar number concentrations and rain water contents to 2M (see Figure 6), and therefore produces similar values for reflectivity (Figure 13j). Above 5 km (which lies just above the simulated freezing level for this case), the differences between the simulations are less pronounced, despite the large variations of ice water content between models at these heights. However, the WB scheme does have slightly higher values of reflectivity than CASIM 1M and 2M. Modifying the parameterization of the sedimentation flux (1M Nr SI) gives results that are more similar to those from the WB scheme.

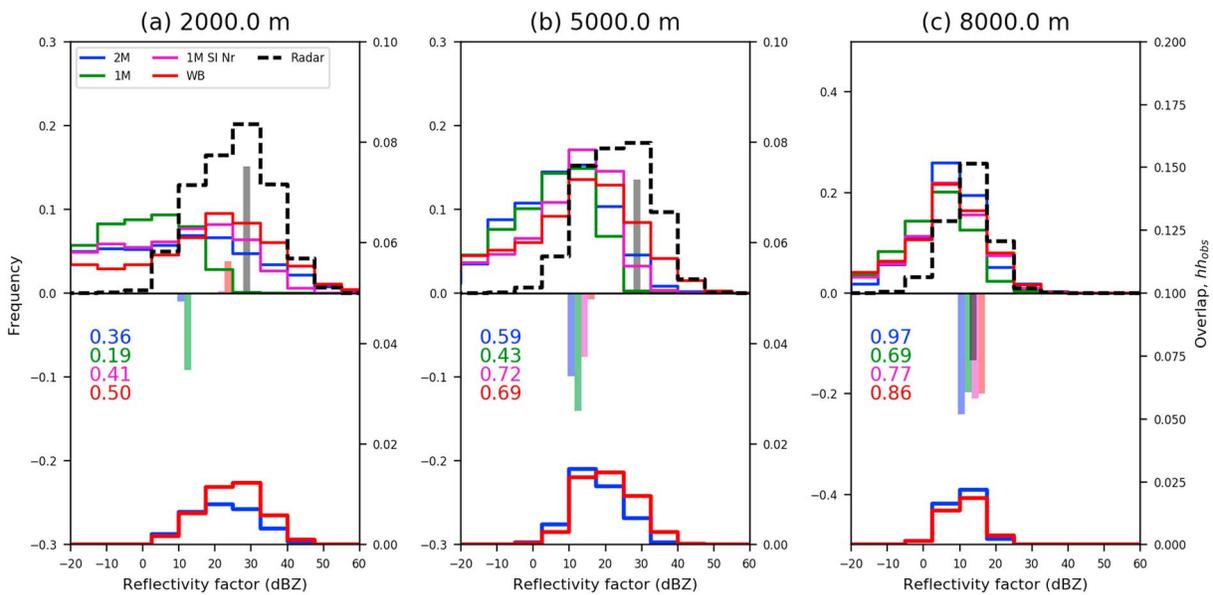
To investigate the effects of microphysics on radar reflectivity, we will use the statistical metrics described in section 2.3 to quantify the differences between the histograms of the simulated and observed reflectivity



**Figure 12.** Histograms of the top-of-atmosphere radiative fluxes from the simulations and the CERES measurements. TOA = top of atmosphere; LW = longwave; SW = shortwave.

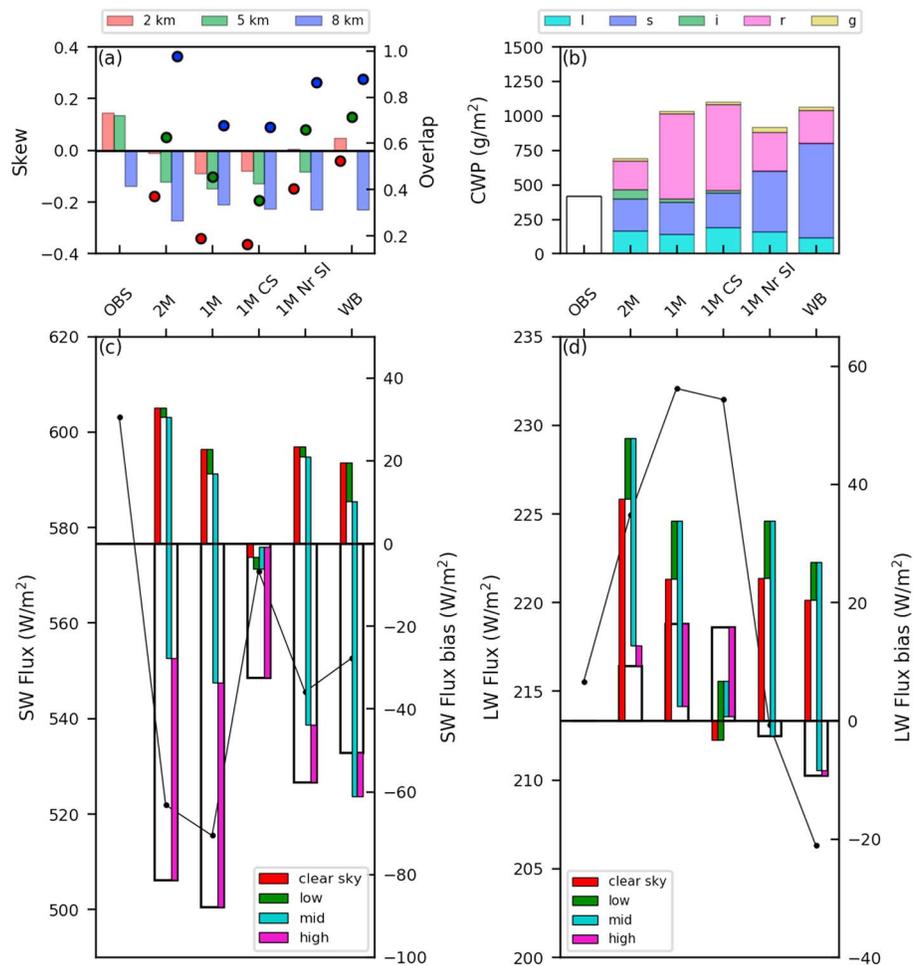


**Figure 13.** Radar reflectivity factors, at 09 UTC on 20 May, in the vicinity of the Guangzhou radar. Panels (a)–(c), show the radar observations interpolated onto fixed heights of 2, 5, and 8 km above mean sea level. The black circles surround the areas which are wholly enclosed within the radar scan volume, at each height. Panels (d)–(o) show the reflectivity simulated by 2M, (d–f); 1M (g–i), 1M Nr SI (j–l), and WB (m–o).



**Figure 14.** Histograms of radar reflectivity and associated statistical metrics at three heights above mean sea level, accumulated between 06 UTC and 18 UTC on 20 May 2017. In the top panels, the lines show the reflectivity histograms (according to the left-hand axes) and the thin vertical bars show the 10- to 30-dBZ skews (heights; right-hand-axes) and median reflectivity bins (horizontal position). The lower panels show the overlap densities,  $hh_{obs}$  for the 2M and WB simulations. The numerical values, printed top-left in the lower panels, give the overlap factors,  $O_h$ , for each of the simulations. The histograms are normalized by the total number of grid points within the radar scan volume at that height; which is the same for the models and (interpolated) radar measurements.

factors. The need for a range of metrics is illustrated in Figure 14, which shows the histograms of radar reflectivity (Figures 14a–14c, upper panels), accumulated between 06 and 18 UTC, within the radar scan volume. Because the radar is insensitive for  $Z < 5$  dBZ, the observed histograms (dashed black lines) are sparsely populated in this range. By contrast, the models can simulate any value of reflectivity. Hence, although the model histograms contain reflectivities below the sensitivity of the radar, these should not be interpreted as evidence of a model bias because clouds particles in this range would be largely invisible to the Guangzhou radar. In particular, when evaluating statistical properties of the plotted histograms, only reflectivities greater than 5 dBZ are used. The heights of the thin vertical bars show the skew parameter,  $s_k$ , for each distribution, and the position of each bar on the horizontal axis indicates the modal reflectivity bin for the corresponding histogram. The lower panels show the overlap densities,  $hh_{obs}$ , for two of the histograms (2M and WB) at each height. The total overlaps,  $\sum hh_{obs} / \sum h_{obs}^2$ , which are measures of the projections of the model histograms onto the observed ones, are printed top-left in each lower panel. We see that no one of the three measures fully describes the errors in the model histograms. For example, at 2 km, 2M and WB (both of which are relatively flat, compared to the observations) have similar values of skew, and adjacent modal bins, but still differ by 30% in terms of overlap with the histogram of the measurements. At 5 km, the same two models overlap to a similar extent with the observations, but WB is relatively symmetric, whereas 2M has large-negative skew. Finally, at 8 km, the differences between the models are less pronounced in all three metrics. Thus, to evaluate which models perform *better* at predicting the statistics of reflectivity (or to quantify intermodel differences), all three metrics are useful. Based on these considerations, Figure 15a compares the skews (bars) and overlaps (circles) of five model configurations to the radar measurements. WB gives the best representation of the observed reflectivities at low levels, although still has significant biases in all three measures. In the mixed-phase layer (5 km), the CASIM 1M and 2M simulations give the worst representations of the measured skew. At this height, the WB scheme has almost no 30–10 dBZ skew, which still compares poorly to the observed positive value of  $s_k$  but represents an improvement over the CASIM simulations. Modifying the ice microphysics in the CASIM scheme (1M Nr SI) gives reflectivity metrics that are more similar to those for the WB scheme. In terms of the overall overlap of the histograms with the observations, the WB and CASIM 2M scheme give the best representations at all heights. For midtropospheric ice clouds (8 km), all the schemes are similar in terms of the three metrics, although 1M and 1M CS stand out as having the lowest overlap with the observed histograms. The 1M and 1M CS schemes are also the worst performing models at low levels, where they give reflectivity histograms with low modes and negative skews that do not overlap well with the measurements.



**Figure 15.** A summary of the main biases found in the 2M, 1M, 1M CS, 1M Nr SI, and WB model configurations: (a) the skews (bars) and overlaps (dots) of the time-averaged reflectivity factor histograms, for the model configurations and radar measurements, at heights of 2 km (red), 5 km (green), and 8 km (blue), above mean sea level; (b) the condensed water paths of each species of hydrometeor for the model configurations and the liquid water path product from MODIS, averaged over the (two) daytime satellite scenes within the case study period; (c, d) the absolute values (black lines) and biases (black and colored bars) of outgoing SW flux, (c), and LW flux, (d), for models and satellite products. The black-outlined bars show the mean flux biases (averaged over two (daytime), and four (day and night), satellite overpass times, for the SW and LW, respectively). The colored bars show the decomposition of the mean flux biases into contributions from each of four cloud types: clear sky (no cloud), red bars; low cloud, green bars; medium (midlevel) cloud, cyan bars; high cloud, purple bars.

The worse performance of 1M can be attributed to the number concentration of rain. Modifying  $N_r$ , gives a reflectivity histogram that is similar to 2M in terms of all three parameters.

### 3.5. Summary of the Effects of Cloud Parameterizations on Model Errors

We can now summarize the main differences between the model configurations in terms of the metrics considered above, and relate these to differences in condensed water contents, number concentrations, and differences in parameterizations. Figure 15b shows the simulated condensed water paths of each species of hydrometeor and the estimate of cloud liquid water path from MODIS. Based on this figure, the models can be divided into groups according to the amount of each species of condensate. Each group has shared characteristic errors compared to the observed quantities (TOA radiation and radar reflectivity):

#### 3.5.1. Rain Water Content

In summary: the 2M, WB, and 1M Nr SI schemes give low RWC simulations; the 1M and 1M CS schemes, high RWC. From this we may conclude that *of the factors considered here, the main factor that influences whether a model produces high or low RWC is the number concentration,  $N_r$ , of raindrops.*

In terms of model errors, the amount of rain is the main factor affecting the radar reflectivity below 5 km. It is weakly correlated with the amount of liquid cloud (possibly due to changes in the collection of cloud by rain) and is of negligible importance for TOA biases. The lower tropospheric radar reflectivities simulated by the high-RWC schemes are too low, compared to the observations. For heights below 5 km, these configurations perform least well, compared to radar observations, because they underpredict the number of large raindrops and therefore have statistical distributions of reflectivity that project poorly onto the measurements. By contrast, in the low-RWC models the mass of rain is distributed across larger sizes of particles, which gives histograms of reflectivity that agree better with the observations.

Although the single- and double-moment CASIM configurations differ in terms of RWC, this difference is related to the fact that the prognostic number concentration in 2M evolves to a lower value than in 1M. The 1M Nr and 1M Nr SI experiments show that reducing the number concentration in the single-moment scheme gives rise to the variations between 1M, on the one hand, and 2M and WB, on the other, in terms of RWC and radar reflectivity.

### 3.5.2. Ice Water Content

The 2M, 1M, and 1M CS schemes give low-IWC simulations. The WB and 1M Nr SI schemes produce higher IWCs. Hence, *of the factors considered here, the main factor influencing the IWC is the parameterization of the sedimentation flux of ice particles.*

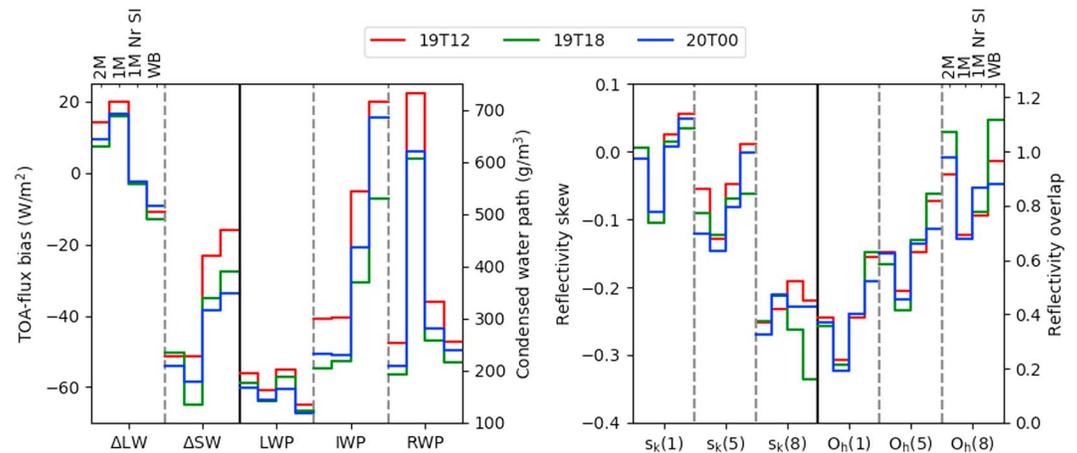
The amount of ice cloud makes a large contribution to the differences in TOA fluxes between the simulations. The low-IWC schemes therefore have positive LW flux biases and the most negative SW flux biases, because they produce relatively tenuous ice clouds that are too transmissive of terrestrial radiation and are not sufficiently reflective of incoming solar radiation. Figures 15c and 15d show the net SW (Figure 15c) and LW biases (Figure 15d), decomposed according to cloud type. The net bias for each model is indicated by the wide black bars, and the thin-colored bars indicate the contributions to the net biases arising from biases in the mean intensities of each cloud type (high, medium, low, and no cloud), according to the metrics described in section 2.3. The contributions from each cloud top type are plotted *end to end* to show how compensating biases between the different cloud types combine to give the overall mean bias. The LW and SW flux biases contributed by high clouds (purple bars) are significantly reduced in the simulations with more ice cloud. Moreover, replacing the sedimentation velocity in 1M with the parameterization from the WB scheme accounts for the differences in the mean flux biases between the CASIM and WB simulations.

The midtropospheric (8 km) radar reflectivities simulated by both the high- and low-IWC schemes are similar to the observations (except at lower values of reflectivity—for which the observations are less reliable, cf. section 2.3). In particular, for all the configurations, the skews of the simulated reflectivity distributions at 8 km are negative and therefore have the same sign as the observed skew. The simulated distributions are, however, more weighted toward lower reflectivities than the observations. The simulated reflectivities are affected by both IWC and the size distribution parameterizations. Hence, the results suggest that the effects of IWC and size distribution on the second moment of the mass distributions is relatively small, for the range of IWCs and temperatures sampled by the simulations.

The differences in ice clouds, between the simulations, are insensitive to the number of microphysical prognostics: the differences between schemes that have one- or two-moment ice, or use separate categories for small crystals and aggregates, are small *provided the schemes use similar microphysical properties*. Hence, the number of prognostic variables has a relatively weak effect on the errors associated with high clouds, in this case. Rather, among the factors tested, the main *influences on* high-cloud errors are the microphysical properties of the ice hydrometeors themselves (ice-particle masses, fallspeeds, and size distributions), assumed in each scheme, *not the number of prognostic variables used*. Moreover, most of the variability can be reproduced by changing a single parameterized process: the sedimentation flux of the ice hydrometeors. The 1M Nr SI experiment shows that modifying the sedimentation flux parameterization to match the one used in WB, give rise to most of the differences between the high- and low-IWC simulations, in terms of IWC, LW flux and SW flux.

### 3.5.3. Liquid Cloud Cover

The schemes, 2M, 1M, WB, etc., with *on-off* cloud fractions ( $RH_c = 1$ ) produce low LWCs; the 1M CS ( $RH_c < 1$ ) scheme, produces high LWCs. Therefore, *of the factors considered, the largest influence on how much liquid cloud*



**Figure 16.** The effects of the forecast initialization time on the sensitivity tests. Each subpanel (as delineated by the vertical dashed lines) shows the variability across the simulations of a different metric. The metric corresponding to each subpanel is indicated along the horizontal axes at the bottom of the plots. The order of the model configurations, within each subpanel, is indicated along the horizontal axis at the top of plot. Each line color corresponds to a different initialization time, according to the legend shown. (a) The variations in TOA flux biases and hydrometeor condensed water paths for each of three initialization times. (b) The variations in skews ( $s_k$ ) and overlaps ( $O_h$ ) of the time-averaged radar reflectivity histograms for each initialization time.

is the parameterization of the critical relative humidity,  $RH_c$ , used for diagnosing cloud water content and liquid cloud fraction and is not actually cloud microphysical in origin.

The amount of liquid cloud in the simulations is the main factor affecting differences in TOA SW flux between the simulations. The low-LWC schemes have large negative SW-flux biases, which are significantly reduced by using a cloud fraction scheme with  $RH_c < 1$ . The decomposition of SW-flux biases, based on the type of cloud (Figure 15c), shows that this bias is due to underestimating the mean intensity of SW radiation reflected from grid columns with low-level and midlevel cloud tops, and overestimating the SW-flux contribution due to columns with no clouds. Using a value of  $RH_c < 1$  increases the area covered by liquid clouds, at the expense of otherwise clear-sky pixels, thereby reducing the net SW-flux bias. Figure 15c also shows that purely cloud microphysical factors have relatively less effect on low-level and midlevel cloud biases. Neither the number of model prognostics nor differences in the assumed properties of cloud and rain hydrometeors have much influence on the low-level and midlevel contributions to the SW and LW biases, compared to the effect of  $RH_c$ .

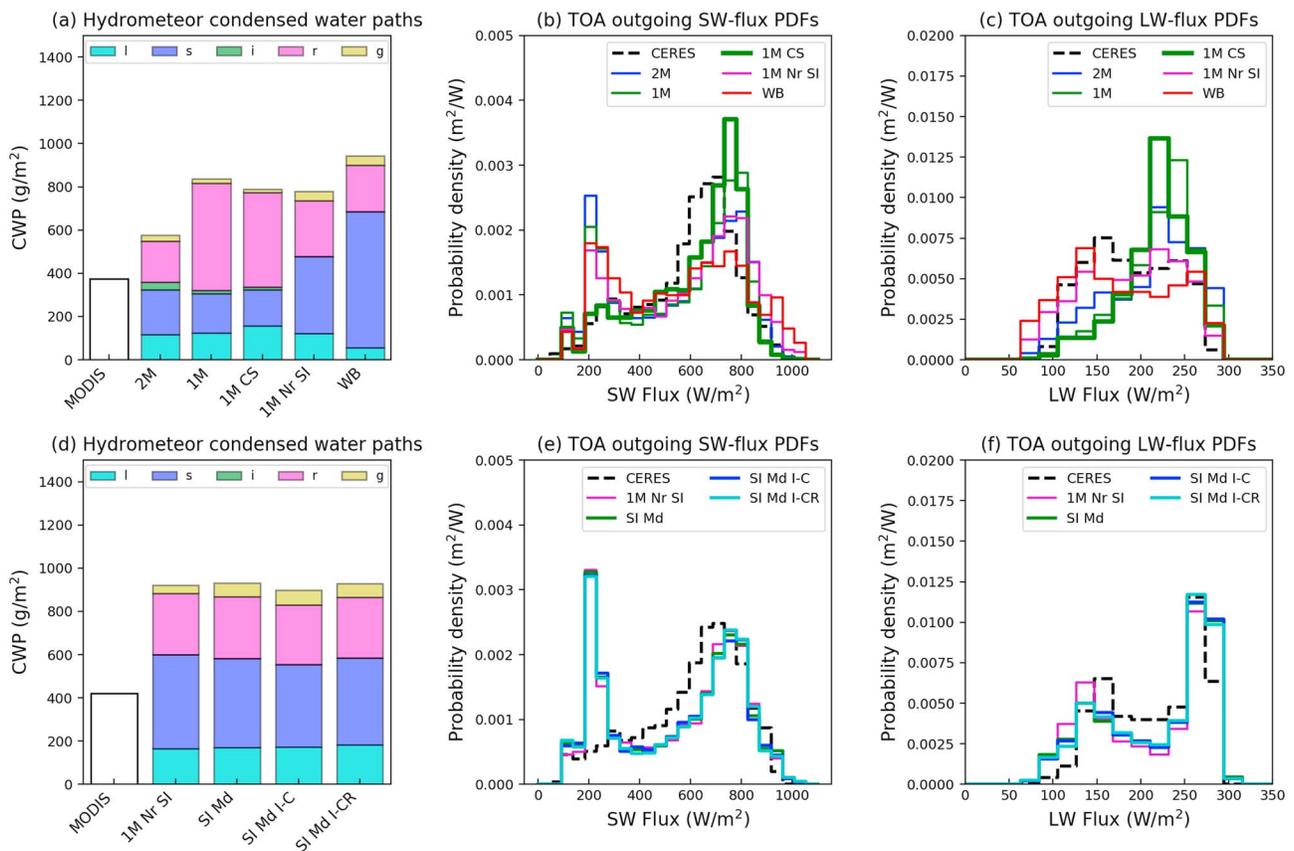
In terms of the influence of  $RH_c$  on radar reflectivity, the small diameters of cloud droplets mean that they mainly affect reflectivities that are below the sensitivity of the radar. Below the melting level, the higher reflectivities are mainly due to rain, the amount of which is not affected much by the amount of liquid cloud.

#### 4. Further Sensitivity Tests

Two limitations of the results presented in section 3 are that we consider only one case study and only a subset of the possible parameterization changes. Hence, we have not explored the sensitivity of the simulations to the meteorological regime, initial conditions, or other interscheme differences in microphysics. To assess the potential impact of these factors on the conclusions, we briefly consider results from a second case study (to assess the impact of large-scale meteorology) and to an ensemble of model initial conditions for the original (SR1) case. Furthermore, for the original (SR1) case, we extend the sensitivity tests to consider some possible additional sources of intermodel variability.

To assess the effects of large-scale variability on the simulations of the SR1 case, we construct a time-lagged ensemble by comparing simulations initialized at 00 UTC 20 May, with simulations initialized 6 and 12 hr earlier. The variability in initial state, across the ensemble members, gives a set of physically realistic perturbations to the evolution of the squall line. This allows us to assess whether the impact of the microphysical parameterization changes is sensitive to the underlying forecasts of the squall line.

Figure 16 compares the radar metrics and satellite metrics for the three initialization times. Each continuous line on the plot shows the variations in a metric, across four different model configurations (2M, 1M, 1M Nr SI, WB).



**Figure 17.** (a–c) The effects of ice microphysical sensitivity tests on the SR2 case study. (a) The condensed water paths of each hydrometeor species; (b, c) the histograms of outgoing SW and LW fluxes at TOA. (d–f) The effects of the additional ice microphysical sensitivity tests on (d) the condensed water paths of each hydrometeor species; (e, f) the histograms of outgoing SW and LW fluxes at TOA. The legend refers to the experiments using the abbreviated nomenclature established in section 4.

and WB), for a given forecast initialization time. It can be seen that the different ensemble members respond in the same way to the parameterization changes. In particular, for all the metrics, selecting a different forecast initialization time (and hence forecast range) would not have altered the conclusions drawn in the proceeding sections. This supports the conclusion that the effects of the parameterizations are generic and not specific to a particular forecast realization.

A possible limitation of using an ensemble of initial conditions to assess the robustness of the parameter sensitivities is that the initial conditions may not be dissimilar enough to give a wide subsample of environmental conditions. To further investigate the dependence of the results on meteorological regime, we select a second case study (hereafter, *SR2*). The second case selected occurred on 12 June 2016. As in the *SR1* case, there was deep convection over South China, but the meteorological situation was less controlled by large-scale frontal boundaries than during *SR1*. Hence, there was no forcing of the convective clouds by thermodynamic gradients and no mesoscale cyclonic features of the kind seen in Figure 2. Instead, local surface-heating-generated convective clouds which were steered eastward by a large-scale westerly flow. Figures 17a–17c show the condensed water paths and TOA-flux histograms for the *SR2* case, averaged over the four satellite overpass times during the simulation period. It can be seen that both the response of the cloud water contents and the subsequent impact on TOA radiation, are similar to those seen for the *SR1* case. In particular, reducing the number concentration of rain decreases the rain water path in the 1M CASIM simulation (Figure 17a). Similarly, allowing  $RH_c$  to be less than one increases the liquid cloud cover, the effect of which can be seen in the SW-flux histograms as a reduction in the SW-flux reflected around  $200 \text{ W/m}^2$ . The effects of the ice microphysical changes are also the same as in *SR1*: the 1M Nr SI experiment shows an increase in IWP, and has TOA-flux histograms, which are similar to those from the WB experiment. In particular, there is an increase on the number of *cold-bright* pixels for which the LW fluxes are in  $50\text{--}150 \text{ W/m}^2$  and SW fluxes are above  $900 \text{ W/m}^2$ .

A further limitation is that only a subset of possible parameterization changes has been considered. In the proceeding sections we have identified a set of parameters which strongly influence the properties of simulated clouds. However, we have not ruled out the existence of other factors which could also have large effects. For example, the above results show that modifying the sedimentation parameterization of ice crystals in the CASIM 1M scheme gives ice clouds which are similar to those in the WB simulation. However, the 1M and WB schemes also differ in other ways. Among these differences, we note the following examples which may influence ice clouds: differences in the PSD parameterizations used in other ice-related processes, for example, collection of liquid by ice crystals or depositional growth of ice; differences in graupel microphysical properties; differences in numerics, for example, the choice of discretization scheme. In this section we will investigate the first of these factors: the use of consistent ice PSD and fallspeed parameterizations across microphysical processes. This is particularly relevant because in the proceeding experiments the WB-PSD and fallspeeds were only implemented in the CASIM sedimentation parameterization, while the other microphysical processes continued to use the original Gamma PSD and CASIM fallspeeds. It is therefore possible that using the same microphysical properties in other ice processes might lead to compensating effects which reduce the resemblance between the WB and modified-CASIM simulations.

Ice microphysical process rates are functions of moments of ice PSD. We can therefore estimate the effects of changing the ice PSD from a Gamma distribution to the generic, WB-parameterization by comparing the relevant PSD moments. Using this approach, an off-line calculation (not shown) suggests that the PSD change will reduce the rate of depositional growth, typically by a factor of between 0.5 and 10 (depending on IWC). Similarly, the effect of the PSD change on riming and ice-rain collisions ranges between a factor of 2 increase and 50% decrease (for IWCs that are typical of the mixed-phase zones in the simulations). These simplified estimates neglect any feedback which may occur, but they motivate the following additional sensitivity tests: (*SI Md*) using the WB-PSD and fallspeeds consistently in sedimentation and depositional growth; (*SI Md I-C*) as experiment *SI Md*, but without collection of cloud-liquid water by ice, that is, without riming; (*SI Md I-CR*) as *SI Md I-C*, but without collisions between ice crystals and raindrops. To assess the overall impact of these experiments, we focus on the metrics presented in Figures 17d–17f which show the effects of the experiments on the masses of each hydrometeor species (Figure 17d) and on histograms of outgoing TOA SW-flux (Figure 17e) and LW-flux (Figure 17f). The radiative-flux histograms are chosen because TOA fluxes are sensitive to small changes in the cloud amounts. It can be seen that the effects of the new sensitivities are much less than the differences between the M1 Nr SI and M1 experiments. In particular, the additional sensitivities give ice water paths which are similar to those from the M1 Nr SI simulation. This lends support to the conclusion that changing the PSD and fallspeed in sedimentation accounts for much of the differences in ice clouds between the 1M and WB simulations. By contrast, the effects of the PSD and fallspeed differences on other ice-related processes are less substantial. An effect of depositional growth rate on LW-fluxes in the intervals 100–150 W/m<sup>2</sup> and 175–225 W/m<sup>2</sup> may be noted, suggesting that some modification of high clouds occurs due to the deposition rate change, but these effects are smaller than those resulting from the sedimentation rate change and will not be investigated further here. A more detailed analysis of the range of metrics considered in section 3 leads to similar conclusions.

## 5. Conclusions

In this paper we have used a case study of heavy rainfall over south China during the presummer rainy season to identify a set of cloud microphysical factors, which are major drivers of errors in the simulation of clouds, precipitation, and radiation in a convection-permitting numerical weather model compared to radar observations and satellite remote sensing. By comparing results from single- and double-moment microphysics schemes, we have evaluated the relative importance of the level of microphysics complexity, that is, the number of model prognostic variables, as compared to the parameterization of individual processes and cloud microphysical properties within each scheme. We have shown that for quantities for which the simulation results show sensitivity to the number of prognostics (primarily the mass of rain, radar reflectivity, and, to some extent, the surface rainfall rate) much of the observed sensitivity can be accounted for by differences in the prognosed and diagnosed number concentrations of rain in the two schemes. Moreover, replacing the diagnosed single-moment number concentration with a value that is more representative of the (lower) concentrations that evolve in the double-moment scheme, gives a single-moment configuration that behaves in a very similar way to the double-moment scheme with respect to rainfall: suggesting that, at least in this case, there is little intrinsic advantage of the computationally more expensive double-moment scheme over

a well-tuned single-moment configuration. This does not imply that the use of double-moment rain adds no value to our simulations, merely that (for this case) the 1M scheme could be made to resemble the 2M scheme by adjusting the diagnosed raindrop number concentration. Moreover, because only one case is studied here, our conclusions do not imply that double-moment schemes do not, in general, confer other advantages. For example, the ability to represent multiple cloud regimes within the same simulation or to resolve interactions between clouds and aerosol. For the East Asian monsoon region, these are questions that we intend to investigate in future work. Neither do we claim that processes outside of those tested here may not exert strong effects on model errors nor that other microphysics schemes and cloud regimes will not give rise to different parameter sensitivities. Indeed, these caveats present interesting topics for future work. For example, we have not investigated how changes to the parameterization of graupel or aerosols might effect the simulations results.

The single- and double-moment realizations of the same microphysics scheme are also shown to be similar in terms of the ice water contents of the clouds produced. Similarly, whether or not the scheme includes separate prognostic variables for small ice crystals and larger ice aggregates is shown to not strongly affect the simulation results. By contrast, the differences in ice clouds (and associated errors in the radiative properties) between two single-moment schemes with widely varying ice microphysical properties (masses, fallspeeds, and particle-size distributions) are shown to be very large. Moreover, these differences—rather than being the complicated result of numerous interacting, process level differences between the schemes—can, to first order, be reproduced almost entirely by the choice of parameterization of a single microphysical process: the sedimentation flux of the ice particles. By *transplanting* the sedimentation flux parameterization from the high-IWC scheme into the low-IWC scheme, the latter can be made to generate ice clouds which closely resemble those produced by the former. Although additional tests to deposition and accretion suggest that these are *second-order* effects (for this case), our results do not imply that other interscheme differences will not interact with fallspeeds to produce compensating or compounding changes. They do serve, however, to highlight the importance of ice crystal sedimentation for intermodel differences and for parameterization development. The sensitivity tests conducted also identify several microphysical properties that contribute strongly to errors in simulated clouds, precipitation, and radiation, for this case study. In particular, errors in the radar reflectivity factors of the simulated clouds, because they are correlated with the mean size of the raindrops, are shown to be strongly influenced by raindrop number. The differences in surface rainfall between single- and double-moment configurations are also shown to arise from the drop numbers; this is particularly the case for the relative *extremes* of rainfall (rain rates in excess of 10 mm/hr). Biases in top-of-atmosphere radiation are sensitive to the amounts of liquid and ice clouds present in the simulations. Hence, ice particle fallspeed, because it controls the amount of ice cloud, is a potentially important influence on the SW errors for high values of the solar reflectance and the LW errors at low values of the terrestrial emittance. Away from regions with deep ice clouds, the main factor controlling radiative biases is the amount of liquid cloud. This was shown to depend only weakly on the choice of microphysics and hence was relatively insensitive to the number of prognostics or to the details of the parameterization of warm-rain processes, or to cloud properties (such as parameterized cloud drop number), which varied between the schemes. Instead, the main factor effecting the amount of liquid cloud was shown to be the choice of critical relative ( $RH_c$ ) in the model's diagnostic cloud fraction scheme. This parameter attempts to model the effects of subgrid-scale variability in water content, and hence is not directly microphysical in origin at all. Although purely microphysical factors (e.g., droplet number and collection efficiencies) must have a role to play in explaining the remaining differences between the schemes, their effect is shown here to be second order compared to the importance of parameterizing the partitioning of water between gas and liquid phases. The results summarized here may be of use for targeting future model developments with the aim of improving forecasts of weather and climate for areas where warm-sector convection plays a major role in the regional climate. In particular, they suggest that in addition to developing more complex microphysics schemes, for example, by increasing the number of prognostic variables, attention can also valuably be paid to optimizing the parameterization of a few key processes in low-order (e.g., single-moment, four-species) schemes. For example, by improving the parameterization of raindrop number concentration, ice particle fallspeed and the variability of relative humidity at unresolved scales (as can be done, on a regional basis, by using ground-based or aircraft measurements), to make cheaper, faster microphysics representations more accurate. In the future, it would be useful to examine the generality of these conclusions by considering further cases, covering a range of weather regimes and types of cloud system, and by comparing results from multiple modeling systems.

**Acknowledgments**

The CERES/Aqua Level 2 Single-Scanner Footprint Edition 3A observed TOA Fluxes can be obtained from the Atmospheric Sciences Data Center at NASA Langley Research Center: <https://ceres-tool.larc.nasa.gov/ord-tool>. The MODIS/Aqua Collection 6 Level 2 Cloud Product data can be obtained from the Level-1 and Atmosphere Archive and Distribution System Distributed Active Archive Center in the Goddard Space Flight Center: [https://ladsweb.modaps.eosdis.nasa.gov/archive/allData/6/MYD06\\_L2](https://ladsweb.modaps.eosdis.nasa.gov/archive/allData/6/MYD06_L2). The reflectivity measurements from the Guangzhou radar and the postprocessed model data can be obtained from the SCMREX data archive: <http://exps.camsma.cn/management/page/staticjspl!PaperData.action>. The Integrated Multisatellite Retrievals for GPM (IMERG) can be obtained from NASA's Precipitation Processing Center: <ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata/>. This work and its contributor Kalli Furtado were supported by the UK-China Research and Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund. Yali Luo and Xi Liu were supported by the Public Welfare Scientific Research Projects in Meteorology (GYHY201406013) and the National Natural Science Foundation of China (91437104 and 41775050).

**References**

Abel, S. J., & Boutle, I. A. (2012). An improved representation of the raindrop size distribution for single-moment microphysics schemes. *Quarterly Journal of the Royal Meteorological Society*, *138*, 2151–2162. <https://doi.org/10.1002/qj.1949>

Baran, A. J., Hill, P., Furtado, K., Field, P., & Manners, J. (2014). A coupled cloud physics-radiation parameterization of the bulk optical properties of cirrus and its impact on the Met Office unified model global atmosphere 5.0 configuration. *Journal of Climate*, *27*, 7725–7752.

Cotton, R. J., Field, P. R., Ulanowski, Z., Kaye, P. H., Hirst, E., Greenaway, R. S., et al. (2012). The effective density of small ice particles obtained from in situ aircraft observations of midlatitude cirrus. *Quarterly Journal of the Royal Meteorological Society*, *139*, 1923–1934.

Ding, Y., & Chan, J. C. L. (2005). The East Asian summer monsoon: An overview. *Meteorology and Atmospheric Physics*, *89*, 117–142.

Doswell, C. A., Brooks, H. E., & Maddox, R. A. (1996). Flash flood forecasting: An ingredients-based methodology. *Weather and Forecasting*, *11*, 560–581. [https://doi.org/10.1175/1520-0434\(1996\)011<0560:FFAIB>2.0.CO;2](https://doi.org/10.1175/1520-0434(1996)011<0560:FFAIB>2.0.CO;2)

Edwards, J. M., & Slingo, A. (1996). Studies with a flexible new radiat ion code. I: Choosing a configuration for a large-scale model. *Quarterly Journal of the Royal Meteorological Society*, *122*, 689–719.

Eidhammer, T., Morrison, H., Mitchell, D., Gettelman, A., & Erfani, E. (2017). Improvements in global climate model microphysics using a consistent representation of ice particle properties. *Journal of Climate*, *30*, 609–629. <https://doi.org/10.1175/JCLI-D-16-0050.1>

Field, P. R., Heymsfield, A. J., & Bansemir, A. (2007). Snow size distribution parameterization for midlatitude and tropical ice clouds. *Journal of the Atmospheric Sciences*, *64*, 4346–4365.

Frey, R. A., Baum, B. A., Menzel, W., Ackerman, S. A., Moeller, C. C., & Spinhirne, J. D. (1999). A comparison of cloud top heights computed from airborne lidar and MAS radiance data using CO<sub>2</sub> slicing. *Journal of Geophysical Research*, *104*(D20), 24,547–24,555. <https://doi.org/10.1029/1999JD900796>

Furtado, K., Field, P. R., Cotton, R., & Baran, A. J. (2015). The sensitivity of simulated high clouds to ice crystal fall speed, shape and size distribution. *Quarterly Journal of the Royal Meteorological Society*, *141*, 1546–1559.

Gettelman, A., & Morrison, H. (2015). Advanced two-moment bulk microphysics for global models. Part I: Off-line tests and comparison with other schemes. *Journal of Climate*, *28*, 1268–1287. <https://doi.org/10.1175/JCLI-D-14-00102.1>

Greenwald, T. J. (2009). A 2 year comparison of AMSR-E and MODIS cloud liquid water path observations. *Geophysical Research Letters*, *36*, L20805. <https://doi.org/10.1029/2009GL040394>

Grosvenor, D. P., Field, P. R., Hill, A. A., & Shipway, B. J. (2017). The relative importance of macrophysical and cloud albedo changes for aerosol-induced radiative effects in closed-cell stratocumulus: Insight from the modelling of a case study. *Atmospheric Chemistry and Physics*, *17*, 5155–5183. <https://doi.org/10.5194/acp-17-5155-2017>

Hilburn, K. A., & Wentz, F. J. (2008). Inter-calibrated passive microwave rain products from the Unified Microwave Ocean Retrieval Algorithm (UMORA). *Journal of Applied Meteorology and Climatology*, *47*, 778–794. <https://doi.org/10.1175/2007JAMC1635.1>

Holloway, C. E., Petch, J. C., Beare, R. J., Bechtold, P., Craig, G. C., Derbyshire, S. H., et al. (2014). Understanding and representing atmospheric convection across scales: Recommendations from the meeting held at Dartington Hall, Devon, UK, 28–30 January 2013. *Atmospheric Science Letters*, *15*, 348–353. <https://doi.org/10.1002/asl2.508>

Holz, R. E., Ackerman, S. A., Nagle, F. W., Frey, R., Dutcher, S., Kuehn, R. E., et al. (2008). Global Moderate Resolution Imaging Spectroradiometer (MODIS) cloud detection and height evaluation using CALIOP. *Journal of Geophysical Research*, *113*, D00A19. <https://doi.org/10.1029/2008JD009837>

Hongru, Y., Huang, J., Minnis, P., Yi, Y., Sun-Mack, S., Wang, T., & Nakajima, T. Y. (2015). Comparison of CERES-MODIS cloud microphysical properties with surface observations over Loess Plateau. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *153*, 65–76.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., et al. (2014). The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, *95*, 701–722.

Huang, S. S. (1986). Heavy rainfall over Southern China in the pre-summer rainy season (in Chinese) (244 pp.). Guangdong Science and Technology Press.

Huang, L., & Luo, Y. (2017). Evaluation of quantitative precipitation forecasts by TIGGE ensembles for south China during the presummer rainy season. *Journal of Geophysical Research: Atmospheres*, *122*, 8494–8516. <https://doi.org/10.1002/2017JD026512>

Huffman, G., Bolvin, D., Braithwaite, D., Hsu, K., Joyce, R., & Xie, P. (2014). Integrated Multi-satellite Retrievals for GPM (IMERG), version 4.4. NASA's Precipitation Processing Center. Retrieved from <ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata/>, accessed 7 April, 2017.

Lakshmanan, V. (2012). Image processing of weather radar reflectivity data: Should it be done in Z or dBZ? *Journal of Severe Storms Meteorology*, *7*(3), 1–8.

Li, Y., Liu, X., & Chen, B. (2006). Cloud type climatology over the Tibetan Plateau: A comparison of ISCCP and MODIS/TERRA measurements with surface observations. *Geophysical Research Letters*, *33*, L17716. <https://doi.org/10.1029/2006GL026890>

Li, P., Zhou, T., & Chen, X. (2017). Water vapor transport for spring persistent rains over southeastern China based on five reanalysis datasets. *Climate Dynamics*. <https://doi.org/10.1007/s00382-017-3680-3>

Lock, A. P., Brown, A. R., Bush, M. R., Martin, G. M., & Smith, R. N. B. (2000). A new boundary-layer mixing scheme. Part I: Scheme description and single-column model tests. *Monthly Weather Review*, *128*, 3187–3199.

Loeb, N. G., Kato, S., Loukachine, K., & Manalo-Smith, N. (2007). Angular distribution models for top-of-atmosphere radiative flux estimation from the clouds and the earth's radiant energy system instrument on the Terra Satellite. Part I: Methodology. *Journal of Atmospheric and Oceanic Technology*, *22*, 338–351.

Loeb, N. G., Kato, S., Loukachine, K., Manalo-Smith, N., & Doelling, D. R. (2007). Angular distribution models for top-of-atmosphere radiative flux estimation from the clouds and the earth's radiant energy system instrument on the Terra Satellite. Part II: Validation. *Journal of Atmospheric and Oceanic Technology*, *24*, 564–584.

Loeb, N. G., Priestley, K. J., Kratz, D. P., Geier, E. B., Green, R. N., Wielicki, B. A., et al. (2001). Determination of unfiltered radiances from the clouds and the Earth's radiant energy system instrument. *Journal of Applied Meteorology*, *40*, 822–835.

Luo, Y., Zhang, R., Wan, Q., Wang, B., Wong, W. K., Hu, Z., et al. (2017). The Southern China Monsoon Rainfall Experiment (SCMREX). *Bulletin of the American Meteorological Society*, *98*, 999–1013.

Luo, Y., Zhang, R., Wan, Q., Wang, B., Wong, W. K., Hu, Z., et al. (2018). The Southern China Monsoon Rainfall Experiment (SCMREX) data archive. Retrieved from <http://exps.camsma.cn/scmrex/management/main!main.action>

Menzel, W. P., Frey, R. A., Zhang, H., Wylie, D. P., Moeller, C. C., Holz, R. E., et al. (2008). MODIS Global cloud-top pressure and amount estimation: Algorithm description and results. *Journal of Applied Meteorology and Climatology*, *47*, 1175–1198.

Miltenberger, A. K., Field, P. R., Hill, A. A., Rosenberg, P., Shipway, B. J., Wilkinson, J. M., et al. (2018). Aerosol-cloud interactions in mixed-phase convective clouds. Part 1: Aerosol perturbations. *Atmospheric Chemistry and Physics*, *18*, 3119–3145. <https://doi.org/10.5194/acp-18-3119-2018>

Morrison, H., & Milbrandt, J. A. (2015). Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I: Scheme description and idealized tests. *Journal of the Atmospheric Sciences*, *72*, 287–311. <https://doi.org/10.1175/JAS-D-14-0065.1>

- Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Monthly Weather Review*, *137*, 991–1007. <https://doi.org/10.1175/2008MWR2556.1>
- Ning, S., Wang, J., Jin, J., & Ishidaira, H. (2016). Assessment of the latest GPM-Era high-resolution satellite precipitation products by comparison with observation gauge data over the Chinese Mainland. *Water*, *8*(11), 481. <https://doi.org/10.3390/w8110481>
- Painemal, D., & Zuidema, P. (2011). Assessment of MODIS cloud effective radius and optical thickness retrievals over the Southeast Pacific with VOCALS-REx in situ measurements. *Journal of Geophysical Research*, *116*, D24206. <https://doi.org/10.1029/2011JD016155>
- Platnick, S., Ackerman, S. A., King, M. D., Meyer, K., Menzel, W. P., Holz, R. E., et al. (2015). MODIS atmosphere L2 cloud product (06\_L2), NASA MODIS Adaptive Processing System, Goddard Space Flight Center, dx.doi.org/10.5067/MODIS/MYD06\_L2.006. *Journal of Atmospheric and Oceanic Technology*, *23*, 395–405. <https://doi.org/10.1175/JTECH1855.1>
- Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., et al. (2017). The MODIS cloud optical and microphysical products: Collection 6 updates and examples from Terra and Aqua. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(1), 502–525.
- Seethala, C., & Horváth, Á. (2010). Global assessment of AMSR-E and MODIS cloud liquid water path retrievals in warm oceanic clouds. *Journal of Geophysical Research*, *115*, D13202. <https://doi.org/10.1029/2009JD012662>
- Sherwood, S. C., Bony, S., & Dufresne, J.-L. (2014). Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, *505*, 37–42. <https://doi.org/10.1038/nature12829>
- Smith, R. N. B. (1990). A scheme for predicting layer clouds and their water-content in a general-circulation model. *Quarterly Journal of the Royal Meteorological Society*, *116*, 435–460.
- Sommeria, G., & Deardorf, J. W. (1977). Subgrid-scale condensation in models of nonprecipitating clouds. *Journal of the Atmospheric Sciences*, *34*, 344–355.
- Stevens, B., & Bony, S. (2013). What are climate models missing? *Science*, *340*(6136), 1053–1054. <https://doi.org/10.1126/science.1237554>
- Tang, X., & Chen, B. (2006). Cloud types associated with the Asian summer monsoons as determined from MODIS/TERRA measurements and a comparison with surface observations. *Geophysical Research Letters*, *33*, L07814. <https://doi.org/10.1029/2006GL026004>
- Van Weverberg, K., Vogelmann, A. M., Lin, W., Luke, E. P., Cialella, A., Minnis, P., et al. (2013). The role of cloud microphysics parameterization in the simulation of mesoscale convective system clouds and precipitation in the tropical Western Pacific. *Journal of the Atmospheric Sciences*, *70*, 1104–1128. <https://doi.org/10.1175/JAS-D-12-0104.1>
- Varble, A., Zipser, E. J., Fridlind, A. M., Zhu, P., Ackerman, A. S., Chaboureaud, J.-P., et al. (2014). Evaluation of cloud-resolving and limited area model intercomparison simulations using TWP-ICE observations. Part 2: Precipitation microphysics. *Journal of Geophysical Research: Atmospheres*, *119*, 13,919–13,945. <https://doi.org/10.1002/2013JD021372>
- Waliser, D. E., Moncrieff, M. W., Burridge, D., Fink, A. H., Gochis, D., Goswami, B. N., et al. (2012). The “Year” of tropical convection (May 2008–April 2010): Climate variability and weather highlights. *Bulletin of the American Meteorological Society*, *93*, 1189–1218. <https://doi.org/10.1175/2011BAMS3095.1>
- Walters, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S., et al. (2017). The Met Office unified model global atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geoscientific Model Development*, *10*, 1487–1520. <https://doi.org/10.5194/gmd-10-1487-2017>
- Wang, R., Chen, J., & Wang, X. (2017). Comparison of IMERG Level-3 and TMPA 3B42V7 in estimating typhoon-related heavy rain. *Water*, *9*(4), 276. <https://doi.org/10.3390/w9040276>
- Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Smith, G. L., & Cooper, J. E. (1996). Clouds and the Earth’s Radiant Energy System (CERES): An Earth observing system experiment. *Bulletin of the American Meteorological Society*, *77*, 853–868.
- Wilkinson, J. M. (2017). A technique for verification of convection-permitting NWP model deterministic forecasts of lightning activity. *Weather and Forecasting*, *32*, 97–115. <https://doi.org/10.1175/WAF-D-16-0106.1>
- Wilson, D. R., & Ballard, S. P. (1999). A microphysically based precipitation scheme for the UK Meteorological Office unified model. *Quarterly Journal of the Royal Meteorological Society*, *125*, 1607–1636.
- Zhang, R. H., Ni, Y. Q., Liu, L. P., Luo, Y., & Wang, Y. (2011). South China Heavy Rainfall Experiments (SCHeREX). *Journal of the Meteorological Society of Japan*, *89A*, 153–166.
- Zhou, X.-J. (2003). Heavy Rainfall Experiment in South China during pre-summer rainy season (HUAMEX), 1998 (in Chinese) (220 pp.). China Meteorological Press.