The Separate Physics and Dynamics Experiment (SPADE) framework for determining resolution awareness: A case study of microphysics

William I. Gustafson Jr.,[1] Po-Lun Ma,[1] Heng Xiao,[1] Balwinder Singh,[1] Philip J. Rasch,[1] and Jerome D. Fast[1]

Received 15 January 2013; revised 18 July 2013; accepted 4 August 2013; published 29 August 2013.

[1] Multiresolution dynamical cores for weather and climate modeling are pushing the atmospheric community toward developing scale aware or, more specifically, resolution aware parameterizations that function properly across a range of grid spacings. Determining resolution dependence of specific model parameterizations is difficult due to resolution dependencies in many model components. This study presents the Separate Physics and Dynamics Experiment (SPADE) framework for isolating resolution dependent behavior of specific parameterizations without conflating resolution dependencies from other portions of the model. To demonstrate SPADE, the resolution dependence of the Morrison microphysics, from the Weather Research and Forecasting model, and the Morrison-Gettelman microphysics, from the Community Atmosphere Model, are compared for grid spacings spanning the cloud modeling gray zone. It is shown that the Morrison scheme has stronger resolution dependence than Morrison-Gettelman, and the partial cloud fraction capability of Morrison-Gettelman is not the primary reason for this difference.

Citation: Gustafson, W. I., Jr., P.-L. Ma, H. Xiao, B. Singh, P. J. Rasch, and J. D. Fast (2013), The Separate Physics and Dynamics Experiment (SPADE) framework for determining resolution awareness: A case study of microphysics, *J. Geophys. Res. Atmos.*, *118*, 9258–9276, doi:10.1002/jgrd.50711.

1. Introduction

[2] There is a growing awareness within the atmospheric modeling community that we need physics parameterizations that work seamlessly across a range of grid spacings [e.g., *Bennartz et al.*, 2011; *Chen et al.*, 2011; *Rauscher et al.*, 2013]. The new crop of climate model dynamical cores, just now becoming available for general use, include the ability to use multiresolution domains with refined grid spacing where necessary and coarser spacing elsewhere. Two examples of these new cores include the Model for Predicting Across Scales (MPAS) [*Skamarock et al.*, 2012] and the High-Order Methods Modeling Environment (HOMME) [*Dennis et al.*, 2005, 2012]. Developments in the mathematics to accurately and efficiently calculate the dynamics on these grids, i.e., the resolved transport and numerical diffusion, have made their use possible [e.g., *Ringler et al.*, 2010], but the handling of physics on theses grids has not kept pace. Until new, so-called *scale aware* or, more specifically, *resolution aware* parameterizations become available, the potential advantages of the multiresolution capabilities will be fettered by the current generation of parameterizations. The models will most likely be able to handle small refinements in grid resolution, but large refinements will introduce excessive error due to differing parameterization behavior.

[3] By being resolution aware, these new parameterizations would either automatically adapt their algorithm to the underlying grid spacing of their host grid column or they would contain techniques that naturally give accurate results across a range of grid spacings [e.g., *Gomes and Chou*, 2010]. With this ideal in mind, one can ask how well current parameterizations meet this goal. However, very little is known in this regard, because testing the resolution dependence of parameterizations is not straightforward. In this study we present a framework we call the Separate Physics and Dynamics Experiment (SPADE) that can be used to isolate the behavior of physics parameterizations across differing resolutions.

[4] Note that the phrases *scale aware* and *scale dependence* are vague. Depending on the context, the term *scale* could refer to how efficiently computational models use computer resources, the size of a particular phenomenon treated by a parameterization, or the size of the model grid spacing used to discretize the atmosphere. All three of these scales are relevant within the context of atmospheric models. Therefore, the remainder of this paper will use the phrases *resolution aware* and *resolution dependence* to refer to how changes in grid spacing impact parameterization behavior. This clearer terminology will prevent confusion, particularly when discussing related issues between

[1]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, Washington, USA.

Corresponding author: W. I. Gustafson Jr., Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, P.O. Box 999, MSIN K9-30 Richland, WA, 99352 USA. (William.Gustafson@pnnl.gov)

©2013. American Geophysical Union. All Rights Reserved.
2169-897X/13/10.1002/jgrd.50711

disciplines, such as atmospheric modelers working with computer scientists to make better climate models.

[5] Past estimates of parameterization resolution dependence have typically been based on one of two approaches. The first is to consider the assumptions that have gone into building the parameterization. For example, the Arakawa-Schubert parameterization assumes every grid box contains an ensemble of clouds, with each cloud being a different height [*Arakawa and Schubert*, 1974]. In comparison, the Kain-Fritsch parameterization uses parcel theory to treat a single representative cloud within the grid column [*Kain and Fritsch*, 1990, 1993; *Kain*, 2004]. Both methodologies are valid, but not necessarily for the same range of grid spacings. The Arakawa-Schubert parameterization is clearly aimed at coarse grid spacings used in climate models, on the order of hundreds of kilometers, while the Kain-Fritsch parameterization is more appropriate for mesoscale models with moderate grid spacings, on the order of tens of kilometers. However, it is not uncommon to see modelers use parameterizations outside of the recommended grid spacing range. There is no clear cutoff where the parameterization suddenly stops working. Instead, the behavior often gradually changes and modelers are lulled into poor parameterization choices by the fact that it is hard to tell when they should no longer use a particular scheme. And the greater use of models in the gray zone of cloud parameterization, around 5–10 km grid spacing, requires some sort of convective treatment, yet there are no schemes that clearly work [*Gerard*, 2007]. So modelers just use the best techniques available even though these are known to be suspect within the gray zone.

[6] The second approach traditionally used to understand resolution dependence is to run a model at different resolutions and then compare the results. Unfortunately, it is very hard to separate and understand the resolution dependence of a particular parameterization from that of the rest of the model. Interactions between schemes and changes in the model dynamics mask changes from the specific parameterization of interest [*Boer and Denis*, 1997]. And in the case of multiresolution models, wave feedbacks from differing parameterization behavior can lead to erroneous wave motions that further complicate the comparison [*Rauscher et al.*, 2013]. Unless strong resolution dependencies exist that can clearly be tracked to specific causes through smartly designed sensitivity tests, e.g., by looking at relative changes in parameterization tendencies during the model spin-up period [*Pope and Stratton*, 2002], one is left to speculate about what causes the resolution-induced differences. The SPADE concept, described in more detail in the next section, is in the vein of this second approach, in that it compares model output from different resolutions. However, SPADE limits the degrees of freedom from the dynamics and selected parameterizations in an attempt to better isolate the resolution dependence of specific parameterizations of interest. In many ways, SPADE is inspired by, and follows from, work by *Williamson* [1999] who examined resolution dependence in the Community Climate Model, version 2 (CCM2) by holding the physics parameterization constant while changing the dynamics resolution. However, SPADE does the opposite, examining resolution dependence by holding dynamics constant while changing the grid spacing of selected parameterizations.

[7] In this first study using the SPADE framework, we have chosen to demonstrate its usefulness by investigating the resolution dependence of microphysics parameterizations, which are alternatively referred to as the stratiform, stratus, or resolved cloud parameterizations in some models. The microphysics parameterizations are responsible for condensing and evaporating clouds, handling phase transitions within clouds, producing precipitation from explicit clouds, and generally preventing supersaturations within grid cells. This is in contrast to the convective parameterizations whose primary purpose is to act as vertical mixers to reduce instability, and that form implicit clouds during the mixing process [*Molinari and Dudek*, 1992]. In models designed for grid spacings on the order of tens of kilometers and smaller, the microphysics normally assumes that everything it does acts over an entire grid cell volume. However, this assumption breaks down for coarser climate models where the microphysics must consider the possibility of partial cloud fractions within a grid cell, appropriately adjusting the volume of the cloud to be less than the total grid cell volume. The determination of which part of the grid cell should contain microphysics-based cloud is explicitly determined by the macrophysics, and this is treated as a separate module in some models such as the Community Atmosphere Model, version 5 (CAM). *But, because the microphysics and macrophysics are so inextricably linked, they are considered as a unit in this study and collectively referred to as the microphysics*. In models such as the Weather Research and Forecasting (WRF) model, where the microphysics typically is assumed to operate over the entire grid cell, there is no explicit concept of macrophysics. However, in reality, it is implicitly assumed that the macrophysics would return only all or no cloud for a given cell, which is the simplest macrophysics assumption possible.

[8] Global and mesoscale modelers have developed methodologies to suit their particular needs. But, as global models begin to approach the resolution of mesoscale models, the appropriate methodology is not always clear. Because coarse (global) and fine (mesoscale) resolution models treat such fundamental concepts such as cloud fraction differently, the resolution dependence of the microphysics is an important issue that needs to be known independently from the other cloud components, such as deep convection. Each cloud parameterization type needs to be examined in isolation, and then once this has been done, the interactions between the parameterizations can be studied to make the entire suite more resolution aware. It is expected that methodologies assuming a binary cloud fraction, i.e., all or no cloud, should exhibit stronger resolution dependence compared to methods allowing for partial cloudiness. However, if the partial cloudiness is determined in an ad hoc manner, it may not respond adequately to resolution changes. Additional differences beyond cloud fraction can also contribute to resolution dependency.

[9] This paper is organized as follows. The next section gives a detailed description of the SPADE framework. Section 3 describes the model configurations for the different simulations used in this study. Section 4 compares the model behavior with observations. Sections 5 and 6 examine the resolution dependence of the Morrison and Morrison-Gettelman microphysics. Section 7 discusses the implications of the results. And section 8 provides a summary.

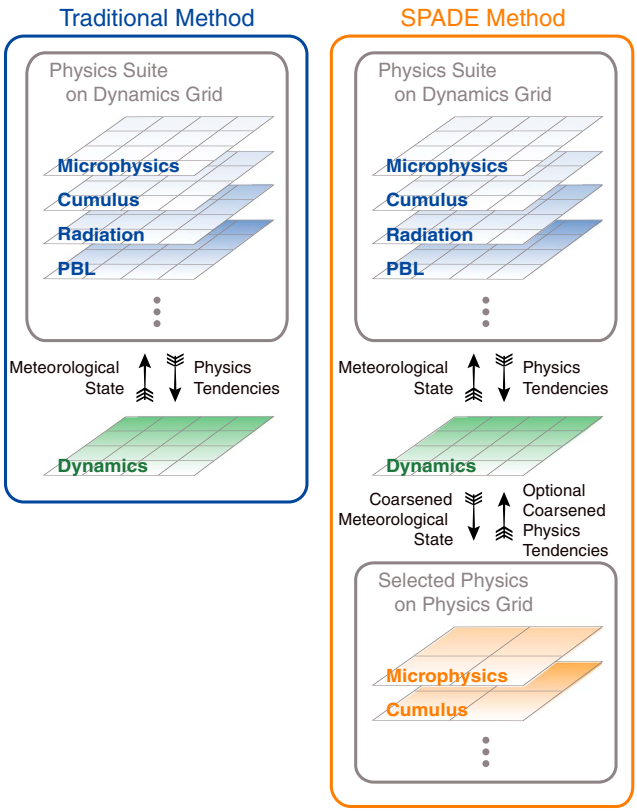

Figure 1. Schematic diagram comparing information flow between traditional and SPADE model simulations. This study does not use the optional coarsened physics tendencies so the results on the dynamics grid are identical between the traditional and SPADE methodologies for a given dynamics grid resolution.

2. The SPADE Framework

[10] As indicated in section 1, there is a need to understand the behavior of physics parameterizations in atmospheric models when used across a range of grid spacings. This study presents a methodology designed around a regional atmospheric model using independent grids for the dynamics and physics portions of the model. Dynamics encompasses advection and numerical diffusion within the model, while physics encompasses the physical processes and subgrid parameterizations such as clouds, radiation, and turbulence. This multigrid capability, which we call the Separate Physics and Dynamics Experiment (SPADE), allows one to easily compare the behavior of specific parameterizations across a range of resolutions while maintaining the same background state for the meteorological variables passed into the parameterizations.

[11] A general schematic of the SPADE concept is shown in Figure 1, which compares the information flow in a traditional model setup with the flow in the SPADE framework. For the analysis presented in this study, the overall concept of SPADE can be pictured as an independent model running at a specified resolution, which then communicates each time step with another copy of the model physics on an alternate grid to determine a second set of physics tendencies and diagnostics. We refer to the first of these grids as the *dynamics grid*, since this is where the transport is done. However, for this study, the model also calculates the physics on the dynamics grid and uses these physics calculations when advancing the model state in time. This results in the output from the dynamics grid being identical to a traditional model run without SPADE. We refer to the alternate grid as the *physics grid*, because its sole purpose is to call physics parameterizations using the alternate grid spacing for analysis purposes—essentially, it produces diagnostic output based on the meteorological state from the dynamics grid. The net result is two sets of output from a single model run: one set of output at the resolution of the dynamics grid that has all the typical output from the model, and a second set of output on the alternate physics grid for the selected physics parameterizations. Because calculations on the physics grid do not alter the dynamics grid for this study, a series of simulations using the same dynamics grid spacing can be performed with various grid spacings on the physics grid to understand how the parameterizations behave when driven by the same meteorological conditions but at different grid spacings. This setup is what is presented in this study, where the dynamics grid is kept at a constant high resolution and the physics grid is coarsened to see how the physics respond. The high-resolution grid is used as the constant meteorological state for comparison between grid spacings, because it is easy to average it to coarser grids and still maintain consistency between variables. If one were to use the coarser grid to feed the high-resolution grid, it would not be possible to add the fine-scale detail that should be present but is unresolved on the coarse grid.

[12] An alternative formulation of SPADE could be used where the physics grid interacts with the dynamics grid such that selected physics tendencies used to advance the model integration come from the physics grid instead of the dynamics grid, which is the methodology used by *Williamson* [1999] with CCM2. This capability is shown in Figure 1 as the "optional coarsened physics tendencies" that can be turned on or off as needed. When using this functionality, the dynamics grid receives feedback from the physics grid, which permits the model to come to an equilibrium state between the behavior of the physics parameterizations at their resolution, with the overall model state defined by the dynamics grid. Work on this type of SPADE setup within WRF has begun, but is beyond the scope of this current paper.

[13] SPADE is currently implemented in the WRF model v3.3.1 [*Skamarock et al.*, 2008]. WRF has been chosen as the host model for several reasons. The first is that, compared to a global climate model, a regional model enables more versatility when designing testing scenarios. Global models usually require processing many input data sets when changing grids and, therefore, typically are intended to run solely at a handful of specific grid spacings. In contrast, regional models are designed to easily move between regions and resolutions. This allows for testing of almost any resolution between cloud scale and global scale grid spacings through the simple process of defining a new grid using the WRF Preprocessing System (WPS). Using a regional model also makes high-resolution tests affordable since global simulations using cloud-resolving grid spacings are prohibitive. By being able

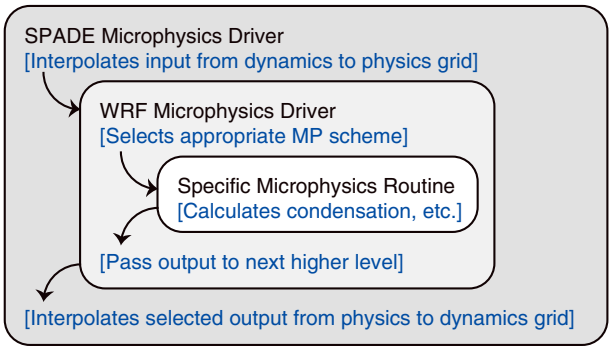

Figure 2. Code flow of the SPADE wrapper around a physics parameterization. Microphysics is shown as an example.

to easily locate the model grid anywhere in the world, different simulations can easily be run where appropriate climate regimes and observations exist.

[14] The second reason for selecting WRF as the host model is that it includes a range of parameterizations based on various assumptions that can serve as a starting point for parameterization tests with SPADE. In addition to the publicly released parameterizations documented in *Skamarock et al*. [2008], the authors have worked as part of a team to port the full suite of physics parameterizations from CAM v5.1 into WRF (P.-L. Ma et al., in preparation, 2013). Initial results from this effort have been released to the community in WRF v3.3 for the modified Zhang-McFarlane deep convection [*Raymond and Blyth*, 1986, 1992; *Richter and Rasch*, 2008; *Zhang and McFarlane*, 1995], University of Washington shallow cumulus [*Park and Bretherton*, 2009], and University of Washington boundary layer [*Bretherton and Park*, 2009] schemes. Since that release, the Morrison-Gettelman microphysics [*Gettelman et al*., 2008; *Morrison and Gettelman*, 2008] and associated macrophysics [*Neale et al*., 2012], and Modal Aerosol Model [*Liu et al*., 2012] schemes have also been ported and recently released in WRF v3.5. These, in combination with the RRTMG longwave and shortwave radiation schemes [*Clough et al*., 2005; *Iacono et al*., 2008; *Mlawer et al*., 1997], already in WRF, provide the full functionality of the CAM atmospheric physics. This allows tests in WRF mimicking the physics suite behavior from the global CAM model.

[15] The third reason for selecting WRF is based on its easily adaptable software framework. Implementation of any new physics parameterization is relatively quick due to the modular nature of the physics layer in the code. Also, many of the tedious tasks required to code an atmospheric model, such as memory management, interprocessor communication, and handling of input/output is handled via WRF's Registry [*Michalakes and Schaffer*, 2004]. The Registry auto-generates thousands of lines of code based on a simple lookup table of variables so that the programmer does not need to worry about the issues listed above. By expanding the subgrid functionality already built into the Registry, which is used by the fire model in WRF [*Coen et al*., 2013; *Mandel et al*., 2011], we have been able to enhance the Registry so it provides the functionality needed for SPADE.

[16] With the enhanced Registry, the other necessary code modification is the ability to map variables between the dynamics and physics grids. This is handled by introducing a SPADE driver module between the initial call to a physics parameterization type and the driver already in WRF. The purpose of the SPADE driver level is to call regridding routines for any required input variables to map them from the dynamics to the physics grid, call the normal WRF physics driver using the modified grid dimensions, and then regrid any output from the WRF physics driver from the physics grid back to the dynamics grid as appropriate. However, most output from SPADE do not need to be returned back to the dynamics grid unless it is being done for a specific purpose, or if the alternate SPADE methodology were to be used where tendencies from the physics grid would be used to advance the model state. An example of the code flow for microphysics is shown in Figure 2. Other physics types follow a similar methodology.

[17] To make horizontal regridding between grids simple and consistent, SPADE requires that the difference between the dynamics and physics grid spacing be an integer multiple. This ensures that the horizontal grid interfaces between grid cells align for the coarsest grid and the ring around the edge of the fine grid's cells that correspond to each coarse grid. This prevents having to split grid cells into pieces. When mapping from the finer to the coarser grid, an average is taken of all the fine grid cells residing within the coarse cell, and that value is applied to the coarse grid cell. This ensures that the mean is maintained between the two grids, so no mass conservation issues are introduced to the model. For the purposes of the present study, there is no need to pass data from the coarser grid back to the finer grid.

[18] In the vertical direction, SPADE assumes that both the dynamics and physics grid reside on the same levels so no vertical regridding is needed. Also, both grids assume the same time step. In theory, one could construct the SPADE driver layers to use a different time step between the grids, but this has not been pursued.

3. Model Configuration

[19] This study presents a series of model simulations corresponding to the time period of the Midlatitude Continental Convective Clouds Experiment (MC3E) field campaign at the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) facility at Ponca City, Oklahoma (http://campaign.arm.gov/mc3e/). This period was chosen because it provides a range of cloud behavior over the central United States ranging from synoptic conditions providing strong large-scale forcing to more quiescent periods where local processes more strongly modulate the cloud behavior. As seen in Figure 3, most of the precipitation fell to the east of SGP.

[20] The model simulations begin on 22 April 2011 12 UTC and extend through 28 May 2011 12 UTC, with the first 24 h excluded from the analyses for spin up. Output is saved hourly for analysis. Data collected during MC3E provides a detailed suite of observations near the center of the model domain for cloud characteristics that are supplemented by additional data sets over the remainder of the domain to ground the simulations in reality. The horizontal model domains used in this study cover a 2016 by 2016 km area from the eastern Rockies to about the Great Lakes in the

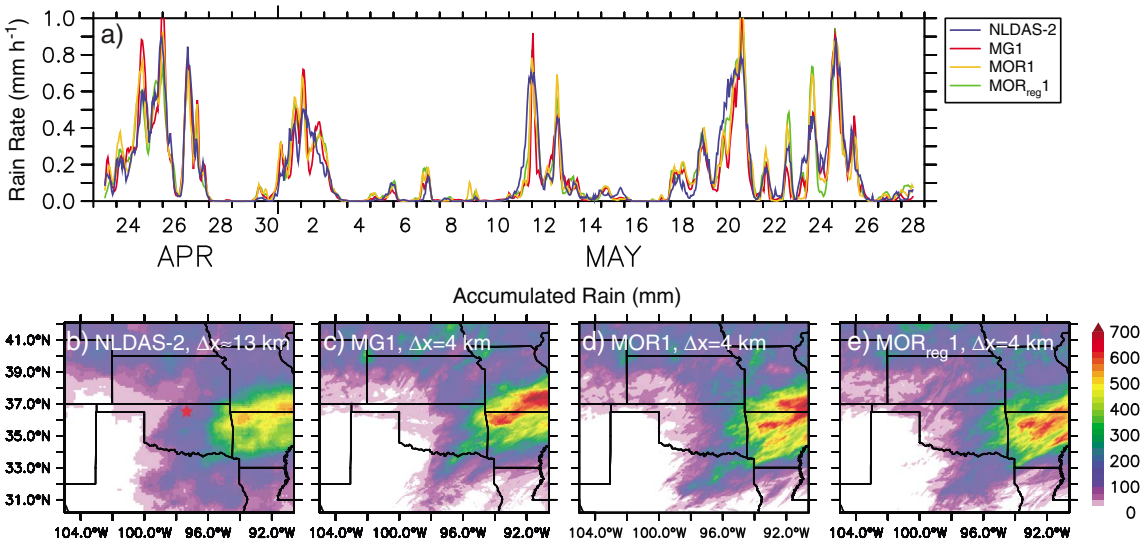

Figure 3. Comparison of NLDAS-2 observations versus modeled precipitation. (a) The domain average precipitation rate (mm h^{-1}), excluding the 160 km lateral boundary relaxation region, for NLDAS-2 observations (blue), MG1 simulation (red), MOR1 simulation (yellow), and $MOR_{reg}1$ (green). Accumulated precipitation (mm) over the period 23 April 2012 12 UTC through 28 May 2012 12 UTC for (b) NLDAS-2, (c) MG1, (d) MOR1, and (e) $MOR_{reg}1$. The red star in Figure 3b indicates the location of the SGP Central Facility.

east-west direction and from the Gulf of Mexico to South Dakota in the north-south direction, with the SGP facility in the exact center of the domain. For reference, the region shown in Figure 3 shows a roughly 12° in latitude by 14° in longitude rectangular section from the interior of the domain.

[21] The primary comparison in this paper is between the Morrison and Morrison-Gettelman microphysics parameterizations. Therefore, the presented simulations are named according to these parameterizations. Two sets of simulations are done using the CAM physics suite in WRF with the microphysics set to Morrison-Gettelman, the MG configuration, or to Morrison for the MOR configuration. Since the Morrison-Gettelman parameterization is well tuned for the CAM physics suite, it could be argued that it might have an advantage over the Morrison scheme, which is normally used with a different set of physics parameterizations in WRF. To determine if this has an impact on the resolution dependence, additional simulations are done using a typical regional climate model physics suite with Morrison microphysics, which is named MOR_{reg}. The specifics for each physics suite are shown in Table 1. When referring to specific simulations from SPADE, a number is appended to the end of the configuration name to identify the grid coarsening factor between the microphysics grid and the grid spacing used for the rest of the model. For example, MG8 refers to the SPADE

Table 1. Model Settings for Each Parameterization Suite

Atmospheric Process	MG/MOR	MOR_{reg}
Longwave radiation	RRTMG-LW	RRTMG-LW
Shortwave radiation	RRTMG-SW	RRTMG-SW
Land surface	Noah	Noah
Surface layer	Monin-Obukhov (Janjic)	Monin-Obukhov (Janjic)
Boundary layer	UW PBL	Mellor-Yamada-Janjic
Deep convection[a]	Zhang-McFarlane	Kain-Fritsch
Shallow convection[b]	UW Shallow Cumulus	Embedded within Kain-Fritsch
Cloud microphysics	Morrison-Gettelman with Park et al. macrophysics (MG) or Morrison (MOR)	Morrison
Trace gases and aerosol	Turned off	Turned off
Tracer/moisture advection	Monotonic	Monotonic
Sixth order diffusion	Turned on at 0.1	Turned on at 0.1

[a]The deep convection schemes are only used for the traditional, 32 km simulations. For the SPADE simulations, deep convection is turned off since the dynamics operate on 4 km grid spacing.

[b]The shallow convection schemes are applied differently between the simulations. For MG and MOR the shallow convection scheme is used for both the traditional 32 km simulation and the 4 km SPADE simulations. For MOR_{reg}, shallow convection is turned on for the traditional 32 km simulation but turned off for the 4 km SPADE simulations. The difference is because the Kain-Fritsch scheme combines both deep and shallow convection into a single parameterization while the CAM suite treats the processes independently, and at 4 km grid spacing some treatment for subgrid cloudiness is still required. Unfortunately, this is not an option with the Kain-Fritsch scheme.

Table 2. Simulations and Regridded Model Output Analyzed in This Study[a]

Physics Configuration	Framework	Grids and Respective Spacing			Type	Run Name
		Dynamics	Physics	Analysis		
Used in Figure 7:						
Morrison-Gettelman w/ CAM suite	Traditional	4 km	N/A	4 km	Simulation	4 km MG
Morrison w/ CAM suite	Traditional	4 km	N/A	4 km	Simulation	4 km MOR
Morrison w/ regional suite	Traditional	4 km	N/A	4 km	Simulation	4 km MOR_{reg}
Morrison-Gettelman w/ CAM suite	Traditional	32 km	N/A	32 km	Simulation	32 km MG
Morrison w/ CAM suite	Traditional	32 km	N/A	32 km	Simulation	32 km MOR
Morrison w/ regional suite	Traditional	32 km	N/A	32 km	Simulation	32 km MOR_{reg}
Morrison-Gettelman w/ CAM suite	Traditional	4 km	N/A	32 km	Regridded	4⇨32 km MG
Morrison w/ CAM suite	Traditional	4 km	N/A	32 km	Regridded	4⇨32 km MOR
Morrison w/ regional suite	Traditional	4 km	N/A	32 km	Regridded	4⇨32 km MOR_{reg}
Used in Figure 8:						
Morrison-Gettelman w/ CAM suite	SPADE	4 km	4 km	4 km	Simulation	MG1
Morrison w/ CAM suite	SPADE	4 km	4 km	4 km	Simulation	MOR1
Morrison w/ regional suite	SPADE	4 km	4 km	4 km	Simulation	$MOR_{reg}1$
Morrison-Gettelman w/ CAM suite	SPADE	4 km	32 km	32 km	Simulation	MG8
Morrison w/ CAM suite	SPADE	4 km	32 km	32 km	Simulation	MOR8
Morrison w/ regional suite	SPADE	4 km	32 km	32 km	Simulation	$MOR_{reg}8$
Morrison-Gettelman w/ CAM suite	SPADE	4 km	4 km	32 km	Regridded	MG1⇨8
Morrison w/ CAM suite	SPADE	4 km	4 km	32 km	Regridded	MOR1⇨8
Morrison w/ regional suite	SPADE	4 km	4 km	32 km	Regridded	$MOR_{reg}1$⇨8
Used in Figure 9:						
Binary cloud frac. Mor.-Gett. w/ CAM suite	SPADE	4 km	4 km	4 km	Simulation	$MG_{BCF}1$
Binary cloud frac. Mor.-Gett. w/ CAM suite	SPADE	4 km	32 km	32 km	Simulation	$MG_{BCF}8$
Binary cloud frac. Mor.-Gett. w/ CAM suite	SPADE	4 km	4 km	32 km	Regridded	$MG_{BCF}1$⇨8

[a]The three grid categories are the dynamics and physics grids from SPADE and the grid used when analyzing the model outputs, e.g., comparing model output at similar resolutions but when the original model simulation is from a higher resolution. The "type" of output is either directly from a simulation or remapped from a simulation onto a coarser grid.

simulation using the CAM physics suite with the Morrison-Gettelman microphysics where the physics grid is eight times coarser than the dynamics grid. Specifically, the side of the grid box is eight times longer leading to an area increase of 64 times. When referring to the general configuration without a specific resolution in mind, the number is excluded from the end of the name.

[22] The choice of grid spacings for this study encompasses the so-called "gray zone" of resolving clouds. This is the transition between scales where the model resolves little of the vertical velocity and other features associated with convective clouds at coarse grid spacings, and where the model resolves most of the convective activity at fine grid spacings. Within this gray zone the model clearly needs some sort of subgrid convective parameterization, yet traditional cloud parameterizations begin to break down because they do not take into account interactions between grid columns for the larger clouds—convective cloud parameterizations typically assume that cloud updrafts only occur within a small portion of the grid column and no information regarding mesoscale organization of convection is communicated between columns to account for compensating motions or semi-resolved features. At the very least, this results in a competition between the resolved and parameterized cloud behavior and a possible double counting of the tendencies for cloud generation. The SPADE grid coarsening factors used in this study are 1 and 8. The dynamics grid for all simulations uses 4 km grid spacing, so the corresponding physics grids have grid spacings of 4 and 32 km. The 4 km grid spacing is roughly the target of global cloud scale resolving models in the next decade and 32 km grid spacing is roughly where the next generation of CAM simulations are beginning to be run. By spanning this range, one gains a better understanding of what to expect when global climate models are pushed to higher resolutions in the coming years.

[23] In addition to the MG8, MOR8, and $MOR_{reg}8$, SPADE simulations based on the 4 km dynamics grid, traditional simulations are performed for the same region using a 32 km grid spacing to compare how the fixed 4 km dynamics constrains the 32 km physics calculations. These additional 32 km simulations are run in the traditional manner with the dynamics and physics on the same grid. These runs are used to demonstrate the limit to which one can attribute differences in model behavior at different scales. To maintain similar boundary forcing between the SPADE runs with 4 km dynamics grid spacing and the runs with 32 km grid spacing, the lateral boundary relaxation region was set to 40 and 5 grid points, respectively. This ensures that the boundary information is applied over the same region, a 160 km ring around the domain edges, for both resolutions.

[24] Between the SPADE and traditional frameworks, a total of 14 simulations are analyzed in this paper, which are listed in Table 2. Additionally, the high-resolution simulations are regridded to the coarse grid to enable resolution consistent comparisons. These coarsened results are indicated by arrows in the run names, e.g., MG1⇨8 indicates regridding of the SPADE results from the 4 km grid to the 32 km grid for the MG physics configuration. The regridding is done using the same algorithm from the SPADE code within WRF, but applied offline to the high-resolution output, so the coarsened high-resolution simulations have the same amount of smoothing as present on the physics grid within the SPADE framework. Note also that while the 4 km traditional and SPADE simulations are listed separately for comparison purposes, they are in fact identical.

[25] Other important model configuration settings include the use of 45 vertical levels, time dependent lateral boundary conditions and sea surface temperatures from the Global Forecast System (GFS) model analyses via the National Climatic Data Center (NCDC), and not using the trace gas and aerosol components of the parameterization suites. Because the aerosols are turned off, one must assume a background aerosol field for the Morrison-Gettelman scheme. This study assumes an aerosol number concentration of 300, 1000, and 0.2 cm^{-3} for the Aitken, accumulation, and coarse modes, respectively for Morrison-Gettelman. The default version of Morrison included with WRF is used for the MOR and MOR_{reg} configurations, so no aerosol assumptions are needed for it. Instead, the Morrison scheme assumes a constant droplet number concentration of 250 cm^{-3}.

4. Comparison With Observations

[26] This section compares the model simulation at 4 km grid spacing, which serves as the basis for the SPADE physics grid calculations, against observations. The point of the comparisons is not to show that the model is perfect. Instead, the goal is to show that the 4 km simulations provide a realistic background meteorological state. The parameterization selections used for the simulations are meant to reflect typical model configurations, and no attempt has been made to fine-tune the configurations specifically for this case. This prevents introducing a bias in the parameterization behavior based on a specialized set of choices. Since the behavior of the microphysics parameterization is the primary focus of the present study, the comparison focuses on the cloud fields.

4.1. Precipitation

[27] Precipitation is the first compared quantity, since it is of primary importance when applying models for climate downscaling and weather forecasting and also is a fundamental output from the microphysics parameterizations. Precipitation from WRF is compared to the North America Land Data Assimilation System Phase 2 (NLDAS-2) product. NLDAS-2 provides hourly precipitation rates on a 0.125° grid by combining daily gauge-based observations, via the National Oceanographic and Atmospheric Administration (NOAA) Climate Prediction Center Parameter-elevation Regressions on Independent Slopes Model (PRISM) product [*Daly et al*., 1994], with hourly National Centers for Environmental Prediction (NCEP) Stage-II Doppler radar data as described in *Mitchell et al*. [2004] and *Xia et al*. [2012].

[28] Averaged over the domain, excluding the lateral boundary relaxation regions, all three 4 km simulations (MG1, MOR1, and $MOR_{reg}1$) capture the overall time dependence of precipitation, as shown in Figure 3a. The early and later portions of the simulated period contain intervals of sustained rain, with quiescent periods interspersed between the two ends. The correlation between the time series of hourly rain rate shown in Figure 3a for NLDAS-2 and WRF is 0.90, 0.91, and 0.93 for MG1, MOR1, and $MOR_{reg}1$, respectively. And the mean bias between the model and NLDAS-2 observations is −0.009, 0.002, and −0.001 mm h^{-1} for MG1, MOR1, and $MOR_{reg}1$, respectively. Based on these statistics, $MOR_{reg}1$ is slightly closer to the observations, but all three generally behave well. None of the model configurations is clearly superior, with one behaving better for some events and another better for other events. So over time, the total precipitation produced by both configurations is similar.

[29] Figures 3b–3e show the spatial patterns of the accumulated precipitation from the three 4 km configurations. The large-scale pattern is similar with lighter rain in the northern half of the domain, no rain in the southwest quadrant, and heavier precipitation in the central eastern portion of the domain. However, each model configuration produces more intense small-scale precipitation than observed, which leads to regions with too much rain offset by other regions with lower rain amounts. While the overproduction of small-scale precipitation by the model configurations relative to observations is most likely real, it should be noted that *Nan et al*. [2010] showed that intense small-scale features are not well captured by NLDAS-2. Also, the model configurations use 4 km grid spacing compared to the 0.125° (~13 km) spacing for NLDAS-2. Both of these issues would lead to more smoothing and lower peak values in NLDAS-2 compared to the two model configurations.

4.2. Cloud Base Height

[30] Cloud base height at the ARM SGP Central Facility is the second compared quantity. Multiple ARM sensors provide estimates of cloud base height, with varying degrees of confidence. The Ka-band ARM Zenith Radar Active Remotely Sensed Clouds Locations (KAZR-ARSCL) value-added product provides a best estimate of cloud boundaries every 4 s using a similar algorithm as the original ARSCL value-added product based on the ARM Millimeter Cloud Radar, which was recently retired [*ARM Climate Research Facility*, 2011; *Clothiaux et al*., 2000]. The Raman lidar also provides an estimate of cloud base height every 10 min, with the cloud base from the "MERGE" product used here [*ARM Climate Research Facility*, 2004]. A third cloud base data set is from the ceilometer at SGP, with readings provided every 16 s [*ARM Climate Research Facility*, 1996]. Figure 4 compares these three estimates of cloud base height with the MG1, MOR1, and $MOR_{reg}1$ simulations. Note that we have identified some of the cloud bases indicated by KAZR-ARSCL as false detections, particularly when the product indicates cloud for very short periods of time in the upper troposphere. To remove these false clouds, data points are only plotted for the KAZR-ARSCL when cloud is indicated continuously for at least 2 min. If any clear sky occurs within a 2 min period, any clouds during that time are not plotted. The resulting impact is to remove noise in the upper troposphere, with very little visual impact at the lower levels. The ceilometer and Raman lidar data help to show when a robust cloud signal is present.

[31] For the most part, the simulated cloud base matches the observed base fairly well when the instruments agree that clouds were present at SGP. As the large-scale conditions change from day-to-day, the model captures the cloud base variability from clouds that form near the surface to those that form in the middle troposphere. Because of natural variability in cloud formation, locally forced clouds not strongly controlled by synoptic conditions do not always form in the model directly over SGP at the same time as in the observations, so a perfect match is not expected. And in fact,

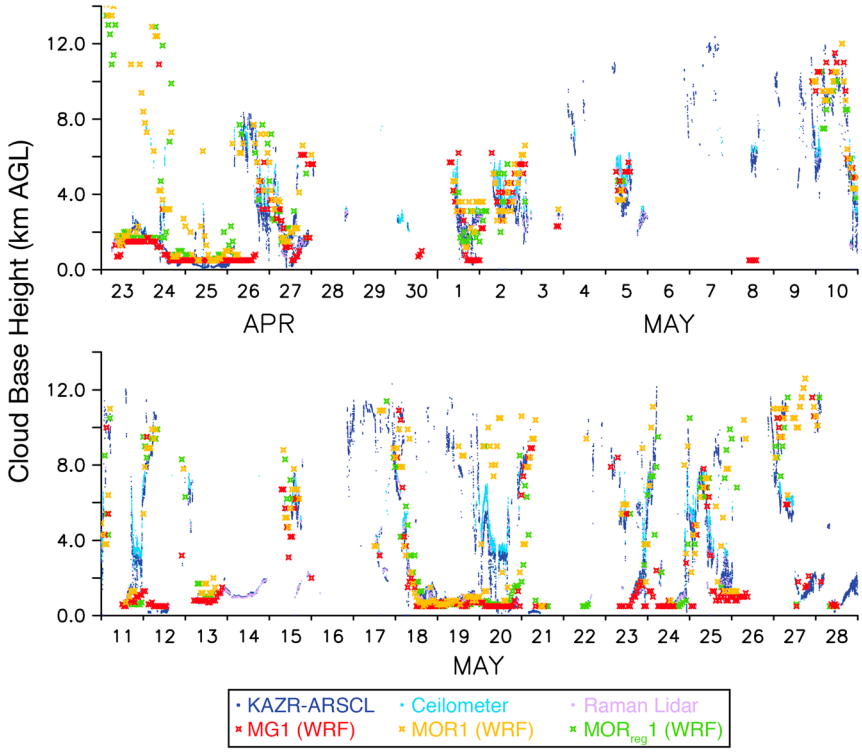

Figure 4. Comparison of observed and simulated cloud base heights at the SGP Central Facility (km above ground level), the location of which is shown in Figure 3b. Observations include the KAZR-ARSCL (dark blue), ceilometer (light blue), and the cloud base height from the merged Raman lidar product (lavender). The simulated cloud base height from MG1, MOR1, and MOR$_{reg}$1 are shown in red, yellow, and green, respectively.

instances occur when the model greatly overestimates or underestimates the cloud base height, e.g., 20 May 2011, but the overall pattern compares well. This is encouraging, especially because this is only a point comparison against a very noisy variable.

[32] The three configurations generate clouds during similar time periods, as expected given the same large-scale forcing for the simulations. However, subtle differences exist between them for the comparison at the SGP Central Facility. MG1 generates cloud mass in a column 49% of the time. By changing the microphysics to Morrison within the CAM suite, this frequency drops to 39% for MOR1. The cloud frequency for MOR$_{reg}$1 is in between, at 41%. This indicates that the physics suite has an influence on how often a microphysics parameterization generates cloud, but the particular microphysics selection also greatly impacts this frequency. For the present cases, the microphysics choice has the strongest effect.

[33] The average cloud base heights are 2.6, 5.0, and 3.9 km above ground level for MG1, MOR1, and MOR$_{reg}$1, respectively. This range can be at least partially explained by fewer occurrences of low clouds in MOR1, which leads to an increased probability for higher-level cloud decks to be identified as the cloud base. For example, early on 24 April and on 20 May, both MG1 and MOR$_{reg}$1 generate clouds below 1 km while MOR1 has a much higher cloud base. The propensity for generating low-level clouds in MG1 is particularly strong compared to the other configurations as can be seen in probability distribution functions (PDFs) shown in Figures 8a, 8e, and 8i, which will be discussed in section 5.2.

4.3. Cloud Radiative Forcing

[34] The last observational comparison moves from the point-measurement at SGP to a domain-scale comparison of top-of-atmosphere (TOA) cloud radiative forcing averaged over the month of May 2011. This represents the net impact of the clouds on the total energy budget as seen at the top of the atmosphere and is an important quantity in climate models for maintaining the long-term consistency of climate simulations. The observations are from the Clouds and Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) cloud radiative effect data set version Ed2.6r [*Loeb et al.*, 2009, 2012] downloaded from http://ceres.larc.nasa.gov/order_data.php. This is a monthly mean product on a 1° by 1° grid, which is primarily based on CERES instruments flown on multiple satellites. For comparison, the simulated cloud radiative forcing is calculated as the difference between the all-sky and clear-sky TOA fluxes from the RRTMG radiation scheme for shortwave and longwave, respectively. Note that because the model lateral boundary conditions are unavailable from NCDC for the last 2 days of May 2011, the model mean is for 2 days less than the observations, from 1 to 28 May versus from 1 to 30 May. This adds a small amount of uncertainty to the comparison, but the overall impression should be the same as would be seen if the model was run for the full month of May.

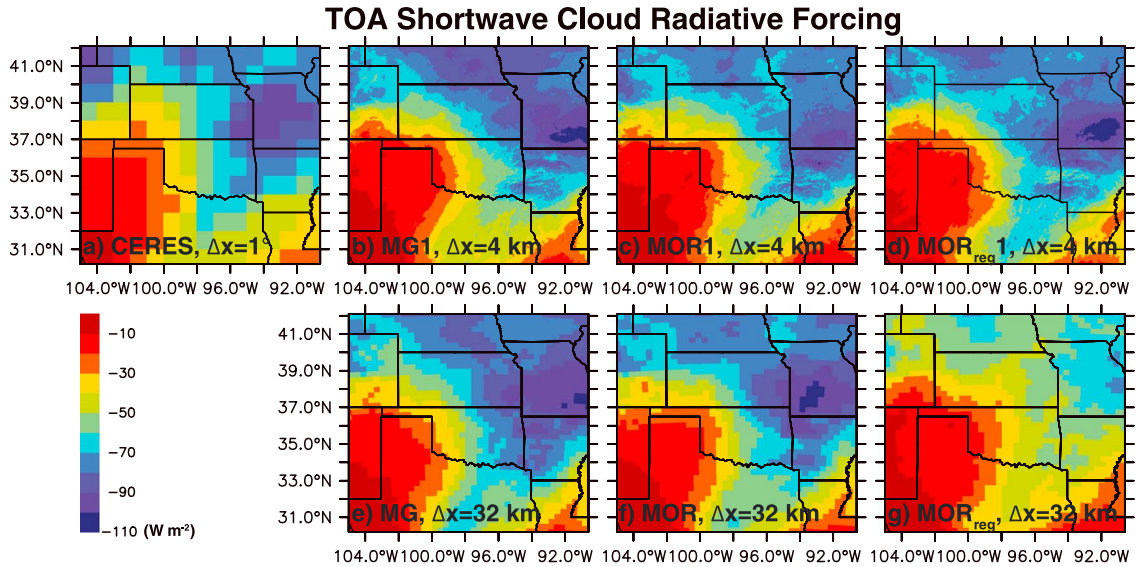

Figure 5. Comparison of observed and simulated top-of-atmosphere shortwave cloud radiative forcing (W m^{-2}) for (a) the satellite based CERES-EBAF product averaged for 1–30 May 2011 versus model results averaged for 1–28 May 2011 for (b) the Δx = 4 km MG1 (c) Δx = 4 km MOR1, (d) Δx = 4 km MOR$_{reg}$1, (e) Δx = 32 km MG, (f) Δx = 32 km MOR, and (g) Δx = 32 MOR$_{reg}$ simulations.

[35] The shortwave cloud radiative forcing comparison is shown in Figure 5 with mean values shown in Table 3. The overall spatial pattern of the forcing is similar to the precipitation pattern shown in Figure 3. The southwest quadrant of the domain has less cloud, and therefore weaker forcing, than the rest of the domain. And the strongest forcing is around the region of high precipitation, near Missouri. The difference in average shortwave cloud forcing between the three 4 km configurations spans 4 W m^{-2} with the weakest forcing from MOR1, the strongest from MOR$_{reg}$1, and MG1 and the CERES observations in between. In this case, changing the supporting physics suite around the microphysics leads to the largest change, but just changing the microphysics within the CAM suite can alter the cloud radiative forcing by over half this amount.

[36] The longwave cloud radiative forcing shown in Figure 6 has similar spatial patterns as the shortwave comparison. However, while the 4 km model simulations compare well against observations for the shortwave forcing, the simulations perform less favorably for the longwave. Both the MOR1 and MOR$_{reg}$1 configurations underestimate the observed CERES value of 27.4 W m^{-2} by about 6 W m^{-2} with MG1 underestimating the forcing by twice this much. For longwave, changing the microphysics has a much stronger impact than changing the supporting physics suite, which is the opposite response from the shortwave comparison.

[37] The frequency and type of cloud occurrence determines the cloud radiative forcing for the different 4 km simulations. One potential contributing factor to the differences is the cloud fraction used for radiation calculations. The average cloud fraction, calculated as the maximum overlap within each column, is 0.44, 0.44, and 0.46, respectively, for MG1, MOR1, and MOR$_{reg}$1. The lower values for the two simulations with the CAM physics contributes to the weaker shortwave forcing, but is not the entire reason. Even though MG1 and MOR1 have the same average cloud fraction, they have very different longwave cloud radiative forcing. The reason for this is the propensity to form clouds with lower water densities with the Morrison-Gettelman parameterization than with Morrison as will be shown in the probability distributions, below (Figures 8a, 8e, 8i). This is clearly evident in profiles of liquid cloud water mixing ratio averaged over cloudy grid cells (not shown). Essentially, MG1 forms clouds with less grid cell mean cloud water than MOR1 and MOR$_{reg}$1. This is at least partly because the Morrison-Gettelman parameterization allows partial cloudiness while Morrison does not—clouds are allowed to form at lower relative humidity when less water vapor is present to condense within a given grid cell. Another feature of the Morrison-Gettelman parameterization is that it forms a larger percentage of the clouds at lower

Table 3. Observed Versus Simulated Cloud Radiative Forcing Values in W m^{-2} With Differences Due to Grid Spacing Shown for Each Traditional Model Configuration[a]

	Shortwave	Longwave
CERES	−51.0	27.4
MG Configuration		
Δx = 32 km	−54.6	14.0
Δx = 4 km	−51.3	15.4
Difference	−3.3 (6%)	−1.4 (−9%)
MOR Configuration		
Δx = 32 km	−52.7	18.2
Δx = 4 km	−48.8	21.7
Difference	−3.9 (8%)	−3.5 (−16%)
MOR$_{reg}$ Configuration		
Δx = 32 km	−37.5	17.8
Δx = 4 km	−52.8	21.5
Difference	15.3 (−29%)	−3.7 (−17%)

[a]Averages are over the area shown in Figure 5 for the period 1–30 May 2011 for CERES and 1–28 May 2011 for the simulations.

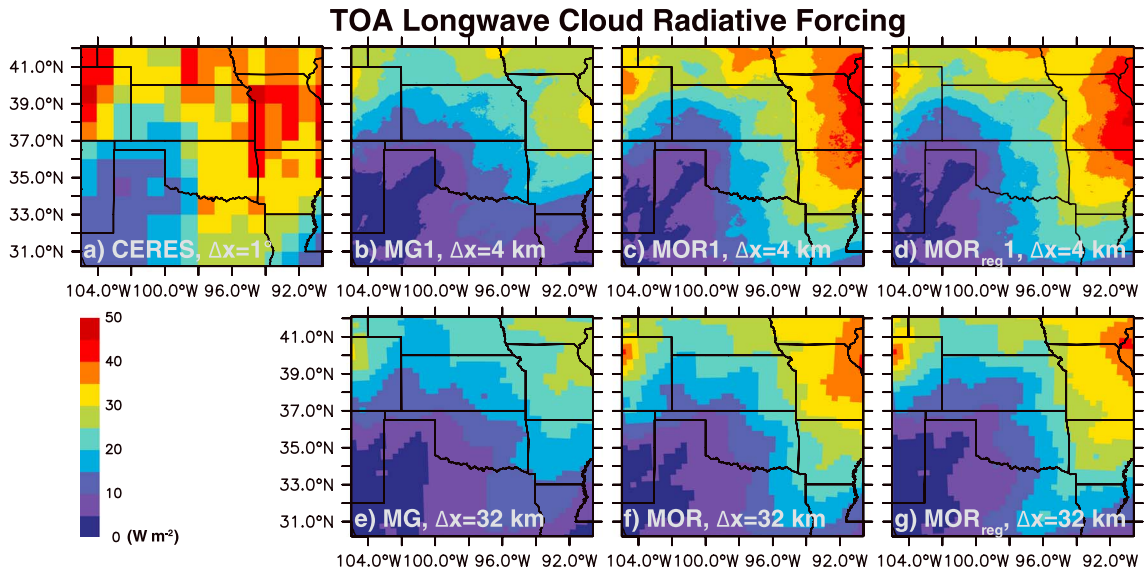

Figure 6. Comparison of observed and simulated top-of-atmosphere longwave cloud radiative forcing (W m^{-2}) for (a) the satellite based CERES-EBAF product averaged from 1–30 May 2011 versus model results averaged for 1–28 May 2011 for (b) the Δx = 4 km MG1 (c) Δx = 4 km MOR1, (d) Δx = 4 km MOR$_{reg}$1, (e) Δx = 32 km MG, (f) Δx = 32 km MOR, and (g) Δx = 32 MOR$_{reg}$ simulations.

levels than the Morrison parameterization (Figures 8a, 8e, 8i). Both the lower densities and lower clouds affect the longwave forcing more strongly than the shortwave, leading to the differences noted above.

5. Resolution Dependence of Microphysics

5.1. Traditional Resolution Comparison

[38] A traditional comparison to identify parameterization behavior at different resolutions is to run a model using two different grid spacings and then compare the results. This works to some extent, in that it shows differences in model behavior at the two resolutions. But this simple comparison has its limits. One cannot fully disentangle which processes lead to the differences. An example of this type of comparison can be found in *Sato et al*. [2008] where a comparison was made using WRF at a range of grid spacings between 3.5 and 28 km. They showed that coarser grid spacings lead to increased error in the diurnal cycle of precipitation over the Tibetan Plateau for their particular model configuration, which used the WRF Single-Moment 6-Class (WSM6) microphysics and no convective scheme for all resolutions. This information is useful, but which parameterization(s) needs to be modified to best improve the behavior across the range of resolutions? Including a convective parameterization might help with 28 km grid spacing, but what about with smaller grid spacings? Sato et al. suggest that coarse grids cannot resolve clouds early in their development when they are small, subsequently leading to erroneous interactions between radiation and surface fluxes, which leads to further errors in precipitation. But, how large is the feedback problem compared with scale issues in the microphysics, such as subgrid variability in the relative humidity, lack of a subgrid macrophysics routine coupled with WSM6, or the lack of proper convective handling at the coarser grid spacings?

[39] Here we further demonstrate the difficulty of using a traditional scale comparison to understand parameterization behavior. As described in section 3, simulations have been performed using the MG, MOR, and MOR$_{reg}$ physics configurations with grid spacings of 4 and 32 km. Comparisons of the TOA cloud radiative forcing between the two grid spacings are shown in Figures 5 and 6 for the shortwave and longwave, respectively, as well as in Table 3. Resolution dependence for these quantities is of particular importance due to the strong role that clouds play in determining climate. And many of the cloud forcing differences due to grid spacing changes would be tuned away within a climate model to maintain the overall climate characteristics. Thus, when a strong resolution dependency exists, it is particularly troublesome.

[40] Looking first at the shortwave cloud radiative forcing in Figure 5 and Table 3, it can be seen that the resolution dependence of the shortwave cloud radiative forcing is on par with the effect of changing the microphysics scheme within the CAM suite, i.e., MG versus MOR, when going from 4 to 32 km grid spacing. The resolution dependence is approximately 3–4 W m^{-2} while the difference between MG and MOR is 2–3 W m^{-2}. When changing the physics suite from CAM to the regional suite, the shortwave cloud radiative forcing for the 32 km MOR$_{reg}$ configuration is an outlier with much weaker forcing, resulting in a very strong resolution dependence for this configuration.

[41] Changing the physics suite, MOR versus MOR$_{reg}$, has a smaller relative impact for the longwave cloud radiative forcing than changing the microphysics scheme, MG versus MOR, which is the opposite behavior from the shortwave forcing. The MG configuration is the outlier for the longwave

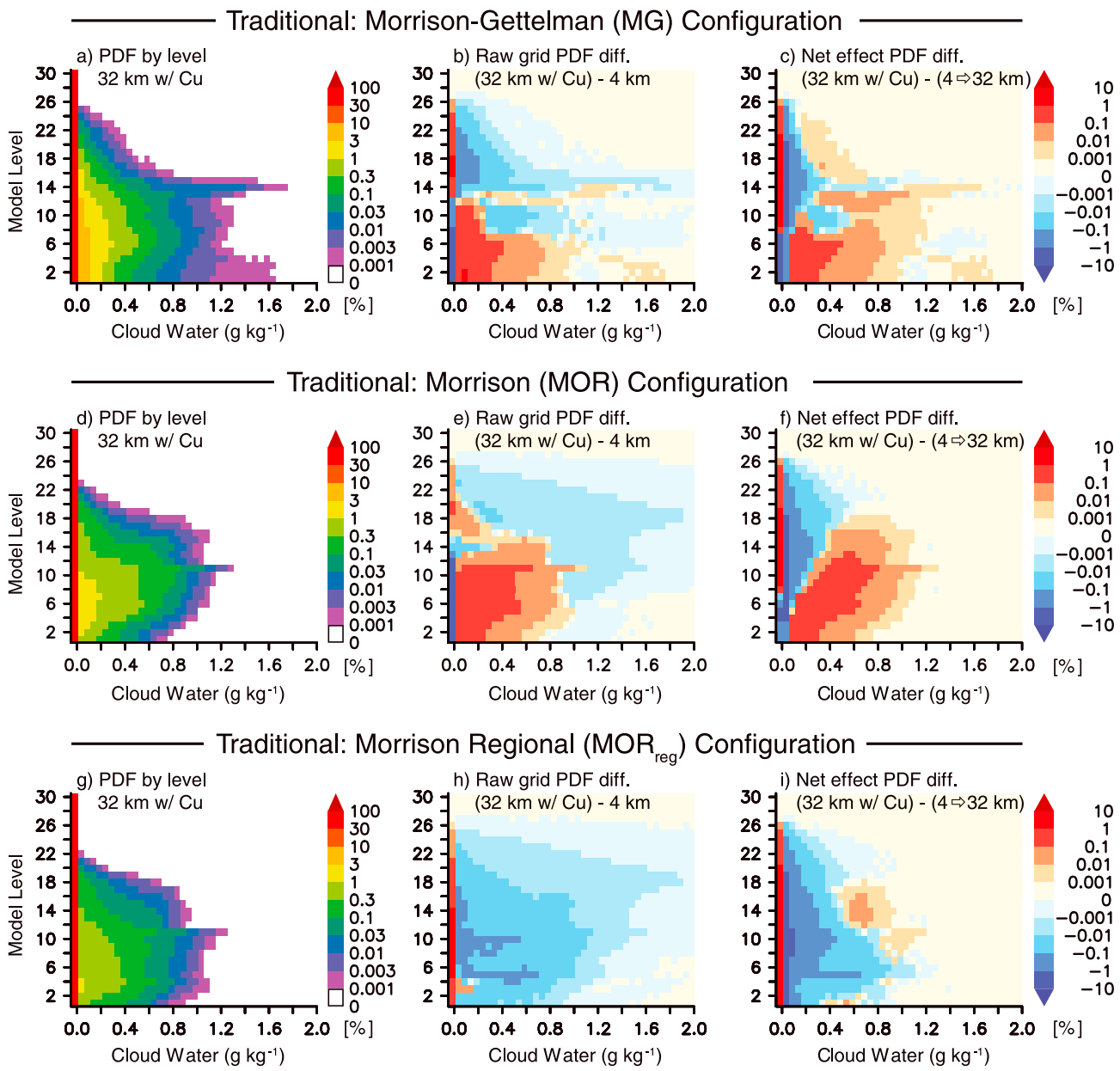

Figure 7. Comparison of probability distributions by model level for grid box mean liquid cloud water mixing ratio for traditional simulations using 32 and 4 km grid spacings. The first column shows the probability distributions for the 32 km simulations using the (a) MG, (d) MOR, and (g) MOR$_{reg}$ configurations. The second column shows the raw grid comparison as the difference in probability between the 32 and 4 km simulations (i.e., 32 km probability minus the 4 km probability) for the (b) MG, (e) MOR, and (h) MOR$_{reg}$ configurations. The third column shows the net effect of the model resolution dependence by comparing the PDFs when calculated on equivalent grids. This is the difference in probability between the 32 km simulation and the 4 km simulation regridded onto a 32 km grid for the (c) MG, (f) MOR, and (i) MOR$_{reg}$ configurations, respectively.

with weaker forcing for both the 4 and 32 km grid spacings. However, MG has a smaller relative resolution dependence, −9%, than the two configurations with Morrison, −16% and −17%.

[42] What resolution induced changes in the clouds lead to MOR$_{reg}$'s larger resolution dependence compared to the MG and MOR configurations? And even when only the microphysics is changed, what leads to the resolution dependence for MG and MOR? Why is the cloud radiative forcing for MOR more resolution dependent than MG? One way to answer these questions is by examining probability distributions of grid box mean liquid cloud water mixing ratio by model level, which are shown in Figure 7. The distributions are calculated using hourly model output from the entire simulation period, excluding the spin-up time, and over all grid points excluding the 160 km relaxation region around

the lateral boundaries. The bin size used to calculate the probabilities is 0.05 g kg^{-1}, and for graphing purposes, the probability of clear-sky grid cells is shown with a negative mixing ratio value as the farthest left column on the plot. A logarithmic color scale is used because most cells do not contain cloud water, and the higher mixing ratio values occur much less frequently. The probabilities are calculated by model level, instead of based on pressure or height in the vertical, to more closely portray how the model represents the clouds within the model grid. For reference, the top of the daytime planetary boundary layer is typically near model level 11 for MOR_{reg} and 9 for MG and MOR, although there is a lot of variability throughout the day depending on the sun's zenith angle and the presence of clouds. The prominent shelf-like feature around model levels 12–14 is the freezing level.

[43] Figures 7a, 7d, and 7g show the grid box mean liquid cloud water mixing ratio probabilities for the MG, MOR, and MOR_{reg} configurations using 32 km grid spacing. It is evident that most cells are clear, and although it cannot be seen from the plots due to the logarithmic scale, the probability exceeds 80–90% for most levels. For cells containing cloud water, the probabilities drop significantly with a monotonic decrease in probability with increasing mixing ratio. While the overall patterns in the probabilities are similar, there are subtle differences. MG concentrates liquid water at lower levels than MOR, has more cells containing low-density clouds, and has a more pronounced increase of dense clouds near the freezing level (around level 14). The mixing ratio probabilities for MOR_{reg} differ from MG similarly to the differences for MOR except that they are even greater. Generally speaking, the clouds generated by the two configurations with the Morrison scheme are more similar to each other than with MG, which uses the Morrison-Gettelman parameterization. So for this context, at 32 km grid spacing, the microphysics selection dominates over the other physics components in determining the mixing ratio characteristics.

[44] Figures 7c, 7f, and 7i show differences in the liquid cloud water probabilities between the 32 and 4 km simulations. Figures 7b, 7e, and 7h present the simplest comparison as a straight difference between the probabilities from the 32 km grid minus those from the 4 km grid. Because the PDFs are calculated on the native grid for each simulation, we call this the raw comparison. This represents the increased or decreased likelihood of finding a particular mixing ratio value if one randomly chooses a grid cell at a particular model level. Red colors indicate an increased probability at the coarser scale compared to the finer scale, while blue colors indicate a decreased probability. It is clear that as one goes from a 4 km grid to a 32 km grid, the three configurations behave differently, particularly MOR_{reg} (Figures 7b, 7e, and 7h). MOR_{reg} generates fewer grid cells containing cloud water at just about every model level with the coarser grid spacing. In contrast, MG has a decreased probability of cells containing cloud water above about level 14, an increased probability of cloud water below level 12, and a transition zone in between. Compared to the base probabilities on the 32 km grid described in the previous paragraph, where the choice of microphysics dominates the type of clouds generated, the resolution dependence is influenced more by the other physics components. The resolution changes for MOR are more like MG than like MOR_{reg}. Even so, there are significant differences between MG and MOR. For example, MOR produces more low-density clouds just above the freezing level. These differences point toward the different resolution dependence for each configuration that needs to be understood.

[45] The raw comparison between the probabilities from the 32 and 4 km grids, described above, is useful for understanding the model behavior when using different resolutions. However, in some ways the comparison is unfair because large grid cells should naturally smooth a noisy field, such as clouds, leading to a reduction in the peak values. How much of the reduced probability of high liquid cloud water content seen in MOR_{reg} and the upper levels of MG and MOR is due to this numerical smoothing effect? One can compensate for this by regridding the 4 km model output onto the 32 km grid and then redoing the probability comparison to see the net effect of the resolution change using equivalent grid spacing. This is shown in the right-hand column of Figures 7c, 7f, and 7i, where the 4 km output has been regridded to the 32 km grid by averaging the 8 × 8 grid points coincident with each corresponding grid point from the 32 km grid, and then subtracted from the 32 km results. This comparison makes it clear that for the grid cells with the largest liquid cloud water content, the difference in grid spacing does bias the results. After this adjustment, the coarser MOR_{reg} simulation actually has a slight increased probability of large liquid cloud water content in the middle levels, and the coarser MG simulation shows an increased probability for the larger liquid cloud water content at most levels. The decreased probability of high cloud water content in MOR is also reduced. Based on these more fair comparisons, if one wanted the 32 km MOR_{reg} simulation to have the same behavior as the 4 km simulation, the 32 km simulation would need to generate fewer midlevel grid cells with high liquid cloud water content and instead spread this water into cells with lower liquid water content. It would also need to reduce the number of clear-sky cells by filling them with low cloud water amounts. Similar behavior would be needed with the 32 km MG simulation for the upper levels, but the lower levels still show the same general behavior as the raw comparison between grids (Figures 7b and 7h). So the lower levels would actually need to generate fewer cells with cloud water on the coarse grid and increase the number of clear cells.

[46] A couple of coherent arguments can be made regarding how the resolution dependency of the liquid cloud water impacts the cloud radiative forcing. The first is that the stronger resolution dependency of the forcing with MOR_{reg} is due to the lower probability of cloud water containing grid cells on the coarser grid. The reduction of cloud water at every model level leaves less cloud water to reflect shortwave radiation back to space, as well as absorb and emit longwave energy. The second argument is that MG and MOR do not have as strong a resolution dependence because even though fewer upper level cells contain cloud water on the coarser grid, this is compensated by an increased probability of cloud water below the freezing level. Essentially, the cloud water forms lower in the atmosphere on the coarser grid.

[47] What can we learn from these grid-dependent differences in model behavior? One obvious point is that the three configurations are not resolution independent since they give different answers depending on the grid spacing. So if these physics suites were used in a multiresolution model, one would get different cloud characteristics depending on the underlying grid spacing, which would generate differing cloud radiative forcing characteristics that would impact climate differently. How can we better isolate the resolution dependence of specific aspects of the model? One way is by removing other strong resolution dependent parameterizations from the system, such as the deep convection, which we tried (not shown). However, turning off the convective parameterization leads to other issues in terms of reduced model accuracy at the coarse resolution. So even though one could show a strong impact by convection, which is the likely culprit for the similar resolution dependence between MG and MOR and a different dependence for MOR_{reg} when making comparisons for the entire model physics suite, it is hard to isolate how to improve a particular scheme using the traditional resolution comparison. By changing one parameterization, interactions with other model components can mask the true resolution dependence of the parameterization being tested. This leads to the more nuanced SPADE methodology presented in the next section.

5.2. SPADE Resolution Comparison

[48] Because of the limitations inherent in the traditional resolution comparison presented above, the SPADE methodology has been developed to better isolate model resolution dependency for specific processes. Here the focus will be on the behavior of the microphysics as an example of how SPADE can be used to better isolate that piece of the cloud parameterization process. Microphysics has been chosen because it is sometimes seen as relatively resolution invariant in that it just reproduces the overall phase transition of water given the saturated or unsaturated conditions within a given grid cell. However, in reality there are subgrid variations in relative humidity, temperature, and cloud structure that could impact the results. By contrasting the microphysics from MG and MOR, it is shown that assumptions made in regional versus global microphysics parameterizations do impact the resolution dependence of the microphysics, and these impacts can be specific to the microphysics and not strongly dependent on the other physics components.

[49] The SPADE methodology is used to isolate the effects of grid cell size changes from the behavior of the parameterization at different scales. Unlike the traditional scale comparison presented above, the SPADE comparison maintains fixed grid spacing for the meteorology, and for the case presented here, the model also uses physics from the same dynamics grid to determine the model state, which is then regridded to the separate physics grid to determine the behavior of the microphysics as an isolated unit. As described in section 3, the base dynamics grid spacing is 4 km, and the coarser scale used for the separate physics grid is a factor of 8 greater at 32 km.

[50] Continuing in the vein of a grid box mean liquid cloud water mixing ratio comparison from the previous section, the left-hand column of Figures 8a, 8e, and 8i, present the probability distribution by level for MG1, MOR1, and $MOR_{reg}1$. For reference, these figures represent the probabilities used for the 4 km simulations in the traditional comparison shown in Figure 7. Similar probability distributions are generated for MG8, MOR8, and $MOR_{reg}8$ (not shown) that are used to calculate the change in probabilities shown in Figures 8b, 8f, and 8j. This second column shows the net change in probability if one were to randomly choose a grid cell at a particular level from a cloud field generated on the 32 km physics grid compared to the 4 km grid. This raw comparison is very similar to Figures 7b, 7e, and 7h, except that the meteorology is now held fixed at a 4 km grid spacing and the resolution difference being compared is solely for the microphysics on the SPADE physics grid.

[51] For this study with SPADE, the 4 km meteorological state drives the physics on both grid spacings, and for the coarser physics grid, the meteorological state is regridded onto the larger grid cells before calling the microphysics routine. This has the advantage of preventing other scale-dependent behaviors from hiding the scale-dependent changes specifically due to the microphysics. A simple way to think about this for scientists familiar with coupled models is that the results in Figure 7 represent an "online coupling" where interactions are allowed between the two components, and Figure 8 represents an "offline coupling" where the interaction is only one way. By comparing Figures 7b, 7e, and 7h and 8b, 8f, and 8j, one can see that resolution induced differences in MOR_{reg} are somewhat similar for the traditional and SPADE scenarios. Both have an increased probability of clear-sky grid cells and a reduced probability of cells with high liquid cloud water content. Where they differ is for cells with low liquid cloud water content. In contrast, the comparison for MG shows much different behavior between the traditional and SPADE scenarios. The opposite behavior for the upper and lower cloud layers in the traditional comparison is gone with SPADE, and it is replaced with a more uniform behavior across all levels. MOR also looses the opposite behavior between upper and lower levels seen in the traditional comparison and the resolution dependence with SPADE is very similar to that seen from MOR_{reg}. So when run in the offline framework of SPADE the Morrison parameterization gives consistent resolution-dependent behavior no matter which physics suite is used, potentially making it easier to diagnose the cause of the dependence.

[52] Of particular interest is that, when isolated from the rest of the model and compared on the raw grids, the resolution dependence of Morrison-Gettelman microphysics appears greater than Morrison. The two parameterizations have an opposite effect on clear-sky probabilities when comparing probabilities from the native grids (leftmost edge of Figure 8b versus Figures 8f and 8j). When isolating the behavior to just the resolution-induced changes within microphysics, one sees that at coarser scales the Morrison (used in MOR and MOR_{reg}) and the Morrison-Gettelman (used in MG) schemes behave differently for clear-sky grid cells. The former leads to a higher probability of clear-sky cells at coarser scales, but the latter leads to the opposite. Instead of completely evaporating clouds away at the coarser scales, as happens in MOR8 and $MOR_{reg}8$, MG8 only partially evaporates the liquid cloud water resulting in a higher probability of grid cells with very low liquid cloud water content, less than about 0.2 g kg^{-1}. Even though these added clouds would be optically thin, they would contribute differently to the cloud radiative

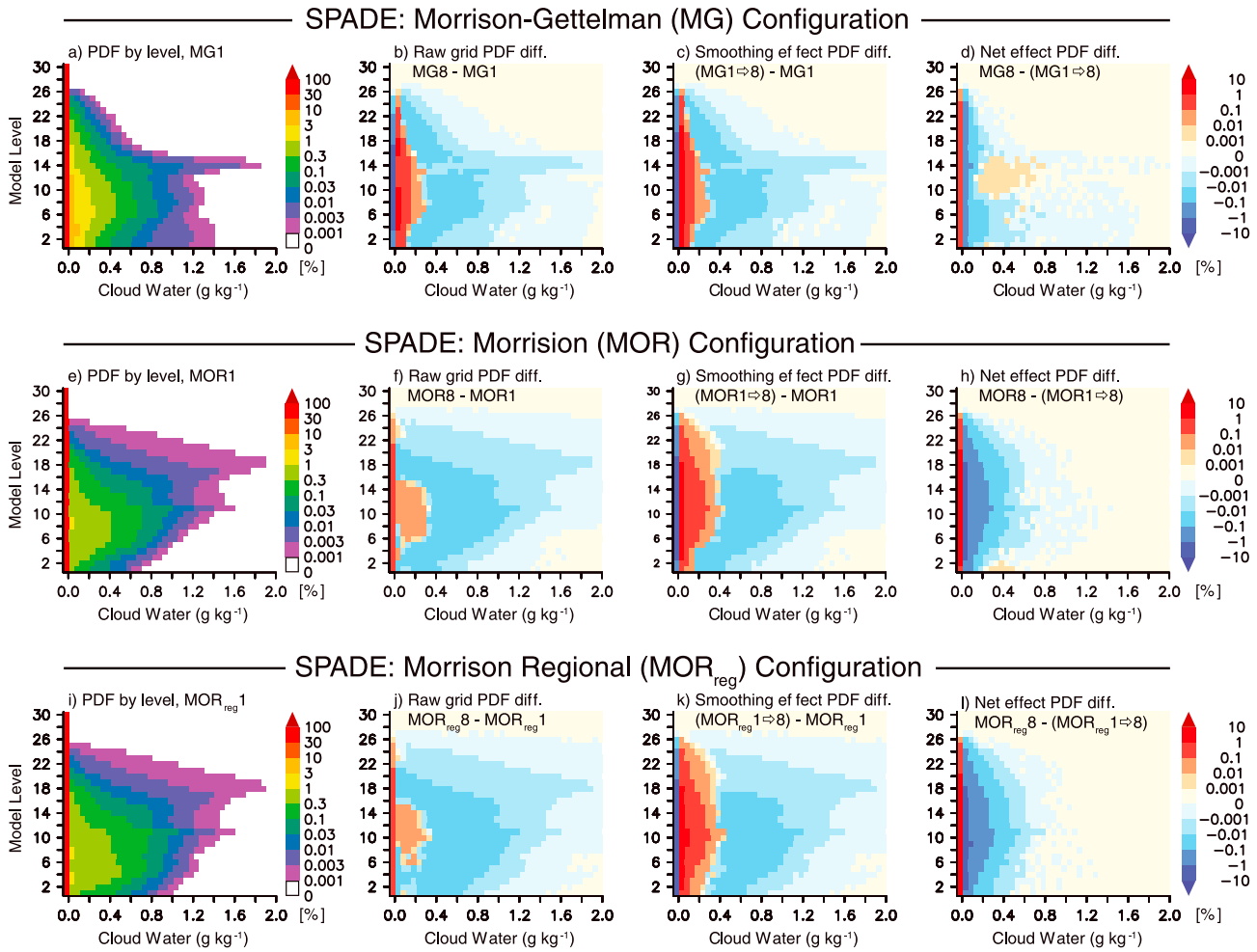

Figure 8. Comparison of probability distributions by model level for grid box mean liquid cloud water mixing ratio for SPADE simulations using a physics-grid coarsening ratio of 8. The first column shows the probability distribution of the liquid cloud water values for (a) MG1, (e) MOR1, and (i) MOR$_{reg}$1. The second column shows the raw grid comparison as the difference between probabilities for the coarse minus the high resolution on the physics grid for the (b) MG, (f) MOR, and (j) MOR$_{reg}$ configurations. The third column shows the change in probabilities due to smoothing from the high-resolution to the coarse grid for the (c) MG, (g) MOR, and (k) MOR$_{reg}$ configurations. The fourth column shows the net effect of the resolution difference as compensated for by the parameterizations. This is the difference in probabilities for the coarse minus the high-resolution grids when compared using an identical 32 km grid spacing for the (d) MG, (h) MOR, and (l) MOR$_{reg}$ configurations.

feedback for a climate model using the MG versus the MOR configuration.

[53] From the raw comparison shown in Figures 8b, 8f, and 8j, one can proceed to separate the impact of smoothing due to larger grid cells from the behavior caused by algorithmic choices in the parameterizations. Figures 8c, 8g, and 8k show the estimated impact solely due to the larger grid cell size in the 32 km grid versus the 4 km grid. This estimate is made by differencing the probabilities for the liquid cloud water from the 4 km grid that has been regridded to the 32 km grid (e.g., MG1⇨8) minus the probabilities from the 4 km grid (MG1). By regridding the 4 km model output to the 32 km grid, one sees what the cloud field should look like on the coarser grid, assuming the 4 km output represents a realistic cloud field. So subtracting the original high-resolution probabilities from this coarsened version of it shows the impact of numerical smoothing on the mean value (for the area of the coarse grid cell) due to changing grid size that must be accounted for by the parameterization if it is to be fully resolution aware. The parameterization at the coarse grid spacing must both generate clouds similarly to the high-resolution clouds and account for this smoothing effect. Ideally, it would also provide statistics of subgrid variability, particularly for fields with strong nonlinear interactions such as cloud fraction. However, at the minimum it must maintain the mean value across resolutions to be considered resolution independent.

[54] The mathematical implications of moving a field to a coarser grid is that high and low values should be smoothed and gradients reduced, with the resulting values pushed toward the mean. This indeed is what the changed probabilities

show in Figures 8c, 8g, and 8k. Because the grid spacing change is identical for all three model configurations, the overall patterns are similar and just show small differences due to different cloud water characteristics on the 4 km grids (Figures 8a, 8e, and 8i). The low values, which in this case are the clear-sky grid cells, become less probable as cloudy cells get smeared into neighboring clear sky. Simultaneously, cells with large cloud water content get smoothed into neighboring cells resulting in lower peak values on the coarser grid. In between, the number of cells with low, but non-zero, cloud water content increases roughly in the range between 0 and 0.4 g kg^{-1}. Because MG1 generates less grid cells on the dense side of the liquid cloud water spectrum, the cutoff point of increased probabilities from the smoothing is slightly lower for MG1 versus MOR1 and $MOR_{reg}1$.

[55] Figures 8d, 8h, and 8l, represent the extent to which the microphysics parameterizations generate the cloud field at equivalent grid spacings, and thus represent a better measure of the resolution awareness for the microphysics. These figures show the net effect of how the parameterization compensates for numerical smoothing and how the algorithm adapts to changes in physical processes at different resolutions. The difference in probabilities shown in Figures 8d, 8h, and 8l is between the physics grid output from the 32 km grid (e.g., MG8) minus the 4 km output that has been regridded to the 32 km grid (e.g., MG1⇨8). Essentially, the raw comparison done on the native grids in Figures 8b, 8f, and 8j has been extended to also remove the difference in grid spacing by removing the smoothing effect shown in Figures 8c, 8g, and 8k. Ideally, the change in probabilities would be zero for Figures 8d, 8h, and 8l if the parameterizations were completely resolution invariant and able to generate the same cloud field at multiple grid spacings. Whether this is the desirable behavior or not is discussed in section 7, but suffice to say, this is not the case. Both the Morrison-Gettelman and Morrison microphysics generate fewer grid cells with cloud water, and consequently too many clear-sky cells, at the coarser scale. It is useful to note how the remaining cells with cloud water differ between the two microphysics parameterizations. Even though MG shows a stronger change between the two grid spacings than MOR when compared on the native grid, Figures 8b and 8f, the differences after compensating for the smoothing are smaller for MG than in MOR, Figures 8d and 8h. This behavior is robust for MOR when changing the physics suite for MOR_{reg}. Something in the Morrison-Gettelman microphysics induces compensating effects for resolution changes, while Morrison does not have this feature and therefore responds more strongly to resolution changes.

6. Continuous Versus Binary Cloud Fraction

[56] What could lead to the greater resolution dependence in MOR versus MG, as shown at the end of the previous section? Even though the microphysics in MOR and MG are both developed based on similar methodologies from Morrison and colleagues [*Gettelman et al.*, 2008; *Morrison and Grabowski*, 2008; *Morrison and Gettelman*, 2008; *Morrison et al.*, 2009], there are very important differences. One is the use of prognostic cloud rainwater in Morrison versus diagnostic cloud rainwater in Morrison-Gettelman. This would imbue a memory between time steps for Morrison that do not exist in Morrison-Gettelman. At smaller grid spacings, this would be much more important since short time steps associated with the small grid spacing would be less than the timescale of the cloud lifetime. For longer time steps, on the order of 30 min, typical of global models, the entire cloud lifetime can occur within a single time step so the memory is less important. However, within the SPADE framework, this issue is not manifested because both grids use an identical time step. An error could be introduced in Morrison-Gettelman because of the short, 15 s time step used with the 4 km grid spacing, but that error should be consistent between MG1 and MG8.

[57] A second difference that could alter the behavior between the two parameterizations is the ability to generate partial stratiform cloud fractions in Morrison-Gettelman versus only a binary cloud fraction in Morrison. The Morrison scheme in MOR uses a binary cloud fraction assumption, with the cloud fraction only taking on one of two values: 0 or 1, or equivalently, clear sky or fully cloudy. The Morrison-Gettelman scheme in MG, along with its associated macrophysics, uses a continuous cloud fraction that can take on a value anywhere between 0 and 1. Barring other differences between the microphysics schemes, such as the gamma PDF assumption regarding in-cloud liquid water mixing ratios in Morrison-Gettelman [*Morrison and Gettelman*, 2008] compared to a monodisperse PDF in Morrison [*Morrison and Grabowski*, 2008], at high resolution the two cloud fraction methodologies should give similar results since each cloud would fill most of a grid cell. However, as the size of the grid cell becomes progressively larger in relation to the average cloud size, the continuous cloud fraction should theoretically give better results than a binary cloud fraction—the continuous cloud fraction should be more resolution aware. Is this the case? The SPADE methodology is a practical way to investigate this question.

[58] To test the hypothesis that the continuous cloud fraction in Morrison-Gettelman gives it greater resolution awareness, we modified the macrophysics in MG to produce a binary cloud fraction instead of the default continuous cloud fraction. Similar to the Morrison scheme, the grid-mean relative humidity is used to determine whether a grid cell is treated as fully clear or cloudy. Liquid cloud fraction is set to 1 when the relative humidity with respect to liquid water within the cell reaches or exceeds 100%, and is set to 0 otherwise. Likewise, the ice cloud fraction is set based on saturation with respect to ice. The condensation/evaporation process for liquid clouds associated with the change of liquid cloud fraction follows the same formulation in MG, and the same maximum overlap assumption used in MG is applied to determine the cloud fraction for mixed phase clouds.

[59] Using this binary cloud fraction methodology for the MG configuration, a new set of SPADE simulations was generated that we identify as $MG_{BCF}1$ and $MG_{BCF}8$ to differentiate from the original MG1 and MG8 simulations. If the continuous cloud fraction is what produces the greater resolution awareness in Morrison-Gettelman compared to Morrison, a comparison between $MG_{BCF}8$ and $MG_{BCF}1$ should appear more like the difference between MOR8 and MOR1.

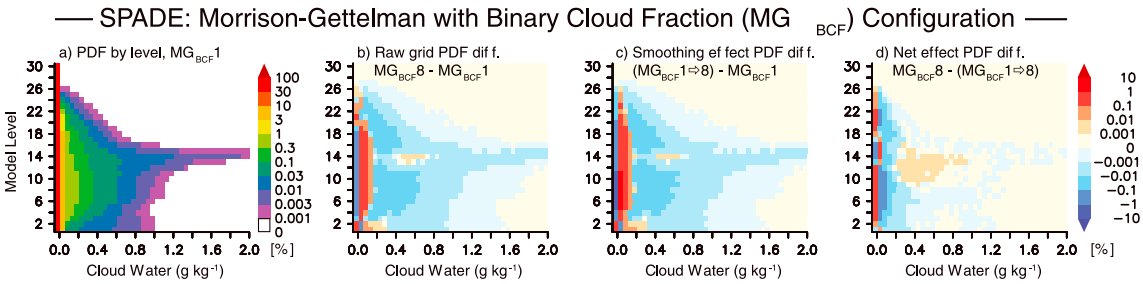

Figure 9. This figure compares the probability distributions by model level for grid box mean liquid cloud water mixing ratio for SPADE simulations similar to the bottom row of Figure 8 with the difference that a binary cloud fraction is imposed within the macrophysics for MG. The stratiform cloud fraction is forced to be one of two values, either clear sky or fully cloudy, for each grid cell.

[60] The results of the binary cloud fraction comparison are shown in Figure 9, which can be compared to the equivalent plots of Figures 8a, 8b, 8c, and 8d that use a continuous cloud fraction. Comparing the left column of each figure shows that the binary cloud fraction limits the amount of cloud water that forms. By not allowing partial cloud fractions, the entire cell must become saturated before cloud water forms with the binary methodology. This means that a higher amount of water vapor is required in a grid cell before cloud water can condense, with the result being a smaller number of cloudy cells. The average stratiform cloud fraction across both clear and cloudy cells drops roughly 50% from 0.41 in MG1 to 0.22 in $MG_{BCF}1$. And the number of occurrences of columns containing stratus cloud drops 68% in $MG_{BCF}1$. After mentally compensating for the tendency toward lower cloud densities for the binary cloud fraction simulations, which draws the probabilities toward the left side of the plots, the raw comparison between the MG_{BCF} runs on their native grids, Figure 9b, are very consistent with the resolution-induced differences using the continuous cloud fraction, Figure 8b. The only notable change in behavior occurs at the few lowest and highest model levels containing liquid cloud water. These layers contain fewer cells with cloud water on the coarse grid when using the binary cloud fraction compared to the rest of the levels that contain more cloud water containing cells. The resolution consistent comparison after adjusting for the smoothing effect of the coarser grid, the net effect shown in Figure 9d, has very similar overall behavior to the continuous cloud fraction simulations (Figure 8d), again with the exception of only a couple levels, which in this case occur around level 14, along with a slight increase in low-density cloudy cells below level 12. This is in stark contrast with the expected behavior if the hypothesis were true; i.e., if use of a continuous cloud fraction was responsible for the greater resolution awareness in Morrison-Gettelman, then the $MG_{BCF}8$ and $MG_{BCF}1$ equivalent grid comparison (Figure 9d) would more closely resemble the MOR8 and MOR1 equivalent grid comparison (Figure 8h), rather than resembling that of the MG8 and MG1 comparison (Figure 8d).

[61] Based on this comparison, we come to the conclusion that the proposed hypothesis is false. For the difference between cloud system and mesoscale grid spacings, the binary cloud fraction is not the primary cause of more resolution invariant behavior of the Morrison-Gettelman microphysics, at least when it operates independently from the rest of the physics suite. Other factors play a stronger role.

[62] The importance of these tests being performed independently from the rest of the physics suite should be highlighted. SPADE allows one to examine the resolution sensitivity of a particular scheme without the complication of interactions with other physics components. For tests that include the interactions, the overall behavior can be very different. These interactions could mask the actual resolution dependency of the scheme being tested, leading to erroneous conclusions. Only after the behavior of each scheme has been examined in isolation can one fully understand how they interact as a suite.

7. Discussion of the Methodology

[63] Several caveats must be kept in mind when using SPADE. The first is that ultimately the development of better parameterizations requires both resolution awareness and improved accuracy. The analysis in this study presents a way to measure the resolution awareness. This also needs to be supplemented with appropriate comparisons against observations. When parameterization developers work on new schemes, they need to determine whether or not the scheme provides sufficient accuracy for the scales intended for its use, and ideally evaluate whether or not parameterization accuracy improves with resolution.

[64] Second, the SPADE concept is based on the assumption that the model reproduces a realistic high-resolution background meteorological state. This is then used to determine the equivalent coarse resolution meteorological state used on the physics grid. For comparisons of microphysics schemes, as done in this study, one assumes that the convective behavior is adequately represented at the high resolution. This requires the selected physics suite to work well at high resolution. Assuming this requirement is met, it permits examination of the microphysics', or any other parameterization type's, resolution awareness in isolation from the other cloud components. It also has the advantage of properly representing the resolution dependence of cloud characteristics that are input into the microphysics. Any resolution-induced errors in the clouds, such as bad behavior from the convective parameterization, are removed from the system by using a coarsened version

of the high-resolution model state from the dynamics grid to drive the physics grid.

[65] Another important caveat of this methodology is that interactions between microphysics and other cloud components are not included when each component is investigated individually. If one wants to design resolution awareness for an entire physics suite, or a subset of more than one parameterization type, then the SPADE methodology would need to be employed across all the schemes of interest. In the present study, the simplest approach is used and only one type, microphysics, has been examined. From here, the next logical step would be to examine the behavior of the microphysics combined with the convective schemes. This will be presented in a future study.

[66] It should be noted that the two microphysics schemes tested in this study were used "out-of-the-box" without any adjustments for the different resolutions. The WRF Morrison scheme is commonly used in the range from a few kilometers to tens of kilometers, so it should be expected to perform respectably at the 4 and 32 km grid spacings used here. However, the Morrison-Gettelman scheme and associated macrophysics were designed for the CAM model, where they typically are used with several hundred kilometer grid spacing, and only recently have excursions into the sub-100 km range become more common. Within the MG configuration, there are several tunable parameters that should be resolution dependent. An example is the relative humidity threshold used to diagnose cloud fraction in the CAM macrophysics. However, exact values are elusive and require extensive tuning to achieve optimal values. So it was decided to use the default settings, which are set for 2° grid spacing in the CAM physics code, to give an idea of how the default parameterization behaves at the mesoscale resolution. Further study to determine which parameter settings give the best compromise between accuracy and resolution awareness is the topic of ongoing and future work.

[67] It is important to stress that establishing resolution awareness requires one to make appropriate choices regarding the scales used for comparison. Inappropriate choices can lead to deceptive results. For example, if one compares the results from MG and MOR based on direct comparison of the 4 and 32km grids, Figures 8b and 8f, respectively, the general resolution dependence of the two microphysics schemes appears opposite of each other for clear-sky conditions. MOR decreases the probability of clouds with higher resolution while MG increases the probability. However, after compensating for the difference in grid size, it is seen that both schemes reduce the probability of clouds with increasing resolution, Figures 8d and 8h.

[68] It is also important to clearly define what is expected of a parameterization in terms of resolution dependence. Whether the difference in resolution dependence identified in this paper between the Morrison and Morrison-Gettelman schemes is good or bad depends on one's perspective. If one seeks a parameterization suite that shows resolution independence, then the knowledge that the WRF Morrison scheme has stronger resolution dependent behavior could lead one to seek a convective parameterization that has equal and opposite behavior. The two could balance each other in the net. This is in fact how some modelers picture the roles of the convective and microphysics parameterizations, which makes physical sense if the convective parameterization is meant to handle subgrid cloud while the microphysics is meant to simulate the resolved cloud. As the grid spacing shrinks, the convective parameterization would be responsible for a smaller portion of the clouds and the microphysics would compensate by having more resolved cloud to reproduce. At mesoscale resolutions this might be true, but at coarser scales this concept breaks down since even stratus clouds can have subgrid horizontal scales. So at spatial scales smaller than convective organization, one may want the microphysics to show resolution dependence, but at scales larger than convective organization, one may want resolution independent microphysical behavior.

[69] As a final discussion topic, conceptually, the resolution awareness of a parameterization requires it to both compensate for the smoothing of fields at coarser scales and the size-induced changes to the particular physical behavior being estimated. The former is solely a numerical issue caused by discretizing the atmosphere onto a grid, while the latter is an issue of what physical processes dominate for different sized regions. If one is designing a resolution independent parameterization, then Figures 8c, 8g, and 8k represent the smoothing that must be overcome by the parameterization. And Figures 8d, 8h, and 8l represent how the algorithmic choices within the parameterization have compensated for the combination of the smoothing plus the size-dependence of the physical process. To the extent that the algorithm can compensate for both processes, the closer the result will be to zero for this column.

8. Summary

[70] The issue of which parameterization to use for a given model grid spacing traditionally has been a function of resolution. Different algorithmic approaches are needed depending on the size relationship between the estimated phenomena and the model grid spacing. In some cases this choice is clear. For example, a large-eddy simulation model does not need a convective parameterization because the convective motions are resolved. This contrasts with a global climate model that requires a detailed handling of subgrid convective motions. Unfortunately, the real atmosphere shows very few clear scale-breaks [e.g., *Lovejoy et al.*, 2010; *Wood and Field*, 2011] so no clear thresholds exist whereby one clearly knows that a particular scheme should be used. The choice is often based on personal or communal experience. And with the advent of multiresolution atmospheric models, this choice becomes even more complicated because the traditional paradigms of parameterization choice at particular grid spacings suggest conflicting choices for different portions of the model grid when the amount of grid refinement is large. In this case, one needs resolution aware parameterizations.

[71] This study presents a new methodology called the Separate Physics and Dynamics Experiment (SPADE) for evaluating the resolution dependence of physics parameterizations in atmospheric models. In theory, it could also be used in any model type with a discretized representation

of fluid flow that requires the use of parameterizations to represent subgrid phenomena, such as ocean models. The SPADE concept uses separate grids for the dynamics and physics portions of the model so they can have independent resolutions. The input from a higher resolution dynamics grid can then be used to drive multiple versions of the physics grid to determine how the physics behavior changes given the same meteorological state that has been regridded to the appropriate spacing of the physics grid. The advantage of SPADE is that it allows one to isolate one or more specific parameterizations to understand the resolution dependence of that piece of the model without conflating the results with resolution dependence from the rest of the model. SPADE also allows one to separate the two issues that must be accounted for by a parameterization when making it resolution aware: the smoothing effect of coarser grids and the effect of changing behavior of the estimated phenomena in proportion to the grid spacing.

[72] As a first demonstration of SPADE, a comparison has been made between a typical mesoscale and a typical global microphysics parameterization. The mesoscale microphysics, the Morrison scheme from WRF, shows a strong resolution dependence based on a comparison of the probabilities of different liquid cloud water concentrations. At coarser scales, the scheme generates lower probabilities of cloudy cells with a monotonically decreasing probability of denser liquid cloud water amounts. In comparison, the Morrison-Gettelman microphysics and accompanying macrophysics from CAM5 have much less resolution dependence for the grid spacings compared, 4 versus 32 km. However, the overall tendency for fewer cloudy cells exists for Morrison-Gettelman as well when compared at the 32 km grid spacing. It was hypothesized that this reduced resolution dependence comes from the partial cloud fraction capability built into the Morrison-Gettelman algorithm. However, using SPADE it was shown that while the partial cloud fraction improves the resolution independence by allowing clouds to form before the whole grid cell becomes saturated, it is not the primary reason for the improvement. Other possible algorithmic differences that could lead to the differences include prognostic versus diagnostic rain and differences in the handling of the ice phase, e.g., the additional graupel phase used in Morrison. The primary reason has yet to be identified and requires further investigation.

[73] **Acknowledgments.** The authors thank Elaine Chapman and Matus Martini for their input on this paper. Funding for SPADE and this paper has been provided by a U.S. Department of Energy (DOE) Early Career grant awarded to William I. Gustafson Jr. Additional funding for the porting of CAM physics into WRF was provided by the PNNL Laboratory Directed Research and Development program and the DOE Office of Science Biological and Environmental Research Program through its Earth System Modeling program. A portion of the research was performed using PNNL Institutional Computing at Pacific Northwest National Laboratory. Data were used from the Southern Great Plains site of the U.S. DOE Atmospheric Radiation Measurement (ARM) Climate Research Facility. NLDAS-2 data used in this study were acquired as part of the mission of NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences Data and Information Services Center. The Pacific Northwest National Laboratory is operated by Battelle Memorial Institute under contract DE-AC05-76RL01830.

References

Arakawa, A., and W. H. Schubert (1974), Interaction of a cumulus cloud ensemble with large-scale environment. Part 1, *J. Atmos. Sci.*, *31*, 674–701, doi:10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2.

ARM Climate Research Facility (1996), updated daily, VCEIL25K. April–May 2011, 36° 36' 18.0" N, 97° 29' 6.0" W: Southern Great Plains Central Facility (C1), cloud base height. Compiled by V. Morris and B. Ermold. Oak Ridge, Tennessee, USA: ARM Data Archive. Data set accessed 26 July 2012, doi:10.5439/1025313.

ARM Climate Research Facility (2004), Updated daily, RLPROF. April–May 2011, 36° 36' 18.0" N, 97° 29' 6.0" W: Southern Great Plains Central Facility (C1), MERGE cloud base height. Compiled by R. Newsom. Oak Ridge, Tennessee, USA: ARM Data Archive. Data set accessed 26 July 2012, doi:10.5439/1027756.

ARM Climate Research Facility (2011), Updated irregularly, ARSCLKAZR1KOLLIAS. April–May 2011, 36° 36' 18.0" N, 97° 29' 6.0" W: Southern Great Plains Central Facility (C1), cloud base best estimate. Compiled by M. Jensen. Oak Ridge, Tennessee, USA: ARM Data Archive. Data set accessed 6 September 2012, doi:10.5439/1052058.

Bennartz, R., A. Lauer, and J. L. Brenguier (2011), Scale-aware integral constraints on autoconversion and accretion in regional and global climate models, *Geophys. Res. Lett.*, *38*, L10809, doi:10.1029/2011gl047618.

Boer, G. J., and B. Denis (1997), Numerical convergence of the dynamics of a GCM, *Climate Dyn.*, *13*, 359–374.

Bretherton, C. S., and S. Park (2009), A new moist turbulence parameterization in the Community Atmosphere Model, *J. Clim.*, *22*, 3,422–3,448, doi:10.1175/2008jcli2556.1.

Chen, Q. S., M. Gunzburger, and T. Ringler (2011), A scale-invariant formulation of the anticipated potential vorticity method, *Mon. Weather Rev.*, *139*, 2,614–2,629, doi:10.1175/mwr-d-10-05004.1.

Clothiaux, E. E., T. P. Ackerman, G. G. Mace, K. P. Moran, R. T. Marchand, M. A. Miller, and B. E. Martner (2000), Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites, *J. Appl. Meteorol.*, *39*, 645–665.

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

Coen, J. L., M. Cameron, J. Michalakes, E. G. Patton, P. J. Riggan, and K. M. Yedinak (2013), WRF-fire: Coupled weather-wildland fire modeling with the weather research and forecasting model, *J. Appl. Meteorol. Climatol.*, *52*, 16–38, doi:10.1175/jamc-d-12-023.1.

Daly, C., R. P. Neilson, and D. L. Phillips (1994), A statistical topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*, 140–158.

Dennis, J., A. Fournier, W. F. Spotz, A. St-Cyr, M. A. Taylor, S. J. Thomas, and H. Tufo (2005), High-resolution mesh convergence properties and parallel efficiency of a spectral element atmospheric dynamical core, *Int. J. High Perf. Comp. App.*, *19*, 225–235, doi:10.1177/1094342005056108.

Dennis, J. M., J. Edwards, K. J. Evans, O. Guba, P. H. Lauritzen, A. A. Mirin, A. St-Cyr, M. A. Taylor, and P. H. Worley (2012), CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model, *Int. J. High Perf. Comp. App.*, *26*, 74–89, doi:10.1177/1094342011428142.

Gerard, L. (2007), An integrated package for subgrid convection, clouds and precipitation compatible with meso-gamma scales, *Quart. J. Roy. Meteor. Soc.*, *133*, 711–730, doi:10.1002/qj.58.

Gettelman, A., H. Morrison, and S. J. Ghan (2008), A new two-moment bulk stratiform cloud microphysics scheme in the Community Atmosphere Model, version 3 (CAM3). Part II: Single-column and global results, *J. Clim.*, *21*, 3,660–3,679, doi:10.1175/2008jcli2116.1.

Gomes, J. L., and S. C. Chou (2010), Dependence of partitioning of model implicit and explicit precipitation on horizontal resolution, *Meteorol. Atmos. Phys.*, *106*, 1–18, doi:10.1007/s00703-009-0050-7.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins (2008), Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, *J. Geophys. Res.*, *113*, D13103, doi:10.1029/2008jd009944.

Kain, J. S. (2004), The Kain-Fritsch convective parameterization: An update, *J. Appl. Meteorol.*, *43*, 170–181, doi:10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2.

Kain, J. S., and J. M. Fritsch (1990), A one-dimensional entraining/detraining plume model and its application in convective parameterization, *J. Atmos. Sci.*, *47*, 2,784–2,802, doi:10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2.

Kain, J. S., and J. M. Fritsch (1993), Convective parameterization for mesoscale models: The Kain-Fritsch scheme, in *The Representation of*

Cumulus Convection in Numerical Models, edited by K. A. Emanuel and D. J. Raymond, pp. 165–170, Meteorological Monographs. Amer. Meteor. Soc., Boston, Massachusetts.

Liu, X., et al. (2012), Toward a minimal representation of aerosols in climate models: Description and evaluation in the Community Atmosphere Model CAM5, *Geosci. Model Devel.*, *5*, 709–739, doi:10.5194/gmd-5-709-2012.

Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong (2009), Toward optimal closure of the Earth's top-of-atmosphere radiation budget, *J. Clim.*, *22*, 748–766, doi:10.1175/2008jcli2637.1.

Loeb, N. G., J. M. Lyman, G. C. Johnson, R. P. Allan, D. R. Doelling, T. Wong, B. J. Soden, and G. L. Stephens (2012), Observed changes in top-of-the-atmosphere radiation and upper-ocean heating consistent within uncertainty, *Nat. Geosci.*, *5*, 110–113, doi:10.1038/ngeo1375.

Lovejoy, S., A. F. Tuck, and D. Schertzer (2010), Horizontal cascade structure of atmospheric fields determined from aircraft data, *J. Geophys. Res.*, *115*, D13105, doi:10.1029/2009jd013353.

Mandel, J., J. D. Beezley, and A. K. Kochanski (2011), Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011, *Geosci. Model Devel.*, *4*, 591–610, doi:10.5194/gmd-4-591-2011.

Michalakes, J., and D. Schaffer (2004), WRF tiger team documentation: The registry, 17 pp., http://www.mmm.ucar.edu/wrf/WG2/software_2.0/registry_schaffer.pdf.

Mitchell, K. E., et al. (2004), The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *J. Geophys. Res.*, *109*, D07S90, doi:10.1029/2003jd003823.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough (1997), Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, *J. Geophys. Res.*, *102*, 16,663–16,682, doi:10.1029/97jd00237.

Molinari, J., and M. Dudek (1992), Parameterization of convective precipitation in mesoscale numerical models: A critical review, *Mon. Weather Rev.*, *120*, 326–344, doi:10.1175/1520-0493(1992)120<0326:pocpim>2.0.co;2.

Morrison, H., and A. Gettelman (2008), A new two-moment bulk stratiform cloud microphysics scheme in the Community Atmosphere Model, version 3 (CAM3). Part I: Description and numerical tests, *J. Clim.*, *21*, 3,642–3,659, doi:10.1175/2008jcli2105.1.

Morrison, H., and W. W. Grabowski (2008), Modeling supersaturation and subgrid-scale mixing with two-moment bulk warm microphysics, *J. Atmos. Sci.*, *65*, 792–812, doi:10.1175/2007jas2374.1.

Morrison, H., G. Thompson, and V. Tatarskii (2009), Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes, *Mon. Wea. Rev.*, *137*, 991–1007, doi:10.1175/2008MWR2556.1.

Nan, Z. T., S. G. Wang, X. Liang, T. E. Adams, W. Teng, and Y. Liang (2010), Analysis of spatial similarities between NEXRAD and NLDAS precipitation data products, *IEEE J. Sel. Top. Appl.Earth Obs. Remote Sens.*, *3*, 371–385, doi:10.1109/jstars.2010.2048418.

Neale, R. B., et al. (2012), Description of the NCAR Community Atmosphere Model (CAM5.0), 274 pp, NCAR Technical Note, NCAR/TN-486+STR, National Center for Atmospheric Research, http://www.cesm.ucar.edu/models/cesm1.0/cam/docs/description/cam5_desc.pdf.

Park, S., and C. S. Bretherton (2009), The University of Washington shallow convection and moist turbulence schemes and their impact on climate simulations with the Community Atmosphere Model, *J. Clim.*, *22*, 3,449–3,469, doi:10.1175/2008jcli2557.1.

Pope, V. D., and R. A. Stratton (2002), The processes governing horizontal resolution sensitivity in a climate model, *Climate Dyn.*, *19*, 211–236, doi:10.1007/s00382-001-0222-8.

Rauscher, S. A., T. D. Ringler, W. C. Skamarock, and A. A. Mirin (2013), Exploring a global multi-resolution modeling approach using aquaplanet simulations, *J. Clim.*, *26*, 2,432–2,452, doi:10.1175/jcli-d-12-00154.1.

Raymond, D. J., and A. M. Blyth (1986), A stochastic mixing model for nonprecipitating cumulus clouds, *J. Atmos. Sci.*, *43*, 2,708–2,718, doi:10.1175/1520-0469(1986)043<2708:asmmfn>2.0.co;2.

Raymond, D. J., and A. M. Blyth (1992), Extension of the stochastic mixing model to cumulonimbus clouds, *J. Atmos. Sci.*, *49*, 1,968–1,983, doi:10.1175/1520-0469(1992)049<1968:eotsmm>2.0.co;2.

Richter, J. H., and P. J. Rasch (2008), Effects of convective momentum transport on the atmospheric circulation in the Community Atmosphere Model, version 3, *J. Clim.*, *21*, 1,487–1,499, doi:10.1175/2007jcli1789.1.

Ringler, T. D., J. Thuburn, J. B. Klemp, and W. C. Skamarock (2010), A unified approach to energy conservation and potential vorticity dynamics for arbitrarily-structured C-grids, *J. Comput. Phys.*, *229*, 3,065–3,090, doi:10.1016/j.jcp.2009.12.007.

Sato, T., T. Yoshikane, M. Satoh, H. Miltra, and H. Fujinami (2008), Resolution dependency of the diurnal cycle of convective clouds over the Tibetan Plateau in a mesoscale model, *J. Meteorol. Soc. Jpn.*, *86*, 17–31.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, W. Wang, and J. G. Powers (2008), A description of the advanced research WRF version 3, 113 pp, NCAR Technical Note, NCAR/TN-475+STR, National Center for Atmospheric Research, http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf.

Skamarock, W. C., J. B. Klemp, M. G. Duda, L. D. Fowler, and S.-H. Park (2012), A multiscale nonhydrostatic atmosperhic model using centroidal Voronoi tesselations and C-grid staggering, *Mon. Weather Rev.*, *140*, 3,090–3,105, doi:10.1175/mwr-d-11-00215.1.

Williamson, D. L. (1999), Convergence of atmospheric simulations with increasing horizontal resolution and fixed forcing scales, *Tellus A*, *51*, 663–673, doi:10.1034/j.1600-0870.1999.00009.x.

Wood, R., and P. R. Field (2011), The distribution of cloud horizontal sizes, *J. Clim.*, *24*, 4,800–4,816, doi:10.1175/2011jcli4056.1.

Xia, Y. L., et al. (2012), Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res.*, *117*, D03109, doi:10.1029/2011jd016048.

Zhang, G. J., and N. A. McFarlane (1995), Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Center general-circulation model, *Atmos.–Ocean*, *33*, 407–446.