

RESEARCH ARTICLE

10.1002/2014JD022245

Special Section:

Fast Physics in Climate Models: Parameterization, Evaluation and Observation

This article is a companion to *Fheng et al.* [2014] doi:10.1002/2014JD022254.

Key Points:

- Produce fine-resolution three-dimensional meteorological fields
- Implement a multiscale data assimilation framework for fine-resolution models
- Improve representations of mesoscale cloud and precipitation systems

Correspondence to:

Z. Li,
Zhijin.Li@jpl.nasa.gov

Citation:

Li, Z., S. Feng, Y. Liu, W. Lin, M. Zhang, T. Toto, A. M. Vogelmann, and S. Endo (2015), Development of fine-resolution analyses and expanded large-scale forcing properties: 1. Methodology and evaluation, *J. Geophys. Res. Atmos.*, 120, doi:10.1002/2014JD022245.

Received 3 JUL 2014

Accepted 24 SEP 2014

Accepted article online 29 SEP 2014

Development of fine-resolution analyses and expanded large-scale forcing properties: 1. Methodology and evaluation

Zhijin Li^{1,2}, Sha Feng^{1,2}, Yangang Liu³, Wuyin Lin³, Minghua Zhang⁴, Tami Toto³, Andrew M. Vogelmann³, and Satoshi Endo³

¹Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA, ²Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, California, USA, ³Environmental Sciences Department, Brookhaven National Laboratory, Upton, New York, USA, ⁴Institute for Terrestrial and Planetary Atmospheres, Stony Brook University, Stony Brook, New York, USA

Abstract We produce fine-resolution, three-dimensional fields of meteorological and other variables for the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) Southern Great Plains site. The Community Gridpoint Statistical Interpolation system is implemented in a multiscale data assimilation (MS-DA) framework that is used within the Weather Research and Forecasting model at a cloud-resolving resolution of 2 km. The MS-DA algorithm uses existing reanalysis products and constrains fine-scale atmospheric properties by assimilating high-resolution observations. A set of experiments show that the data assimilation analysis realistically reproduces the intensity, structure, and time evolution of clouds and precipitation associated with a mesoscale convective system. Evaluations also show that the large-scale forcing derived from the fine-resolution analysis has an overall accuracy comparable to the existing ARM operational product. For enhanced applications, the fine-resolution fields are used to characterize the contribution of subgrid variability to the large-scale forcing and to derive hydrometeor forcing, which are presented in companion papers.

1. Introduction

The Earth's climate system involves a variety of physical processes that span a wide range of spatial and temporal scales. These processes fundamentally influence climate and climate change, but often occur on scales that are too small for typical global climate models (GCMs) to resolve; so, these unresolved physical processes must be parameterized in such models. Aerosol, cloud, and precipitation processes and their interactions that are known as "fast physics" are among these processes, and their parameterizations have remained one of the greatest sources of uncertainty in climate models, as explained in the well-known "Charney Report" [Charney *et al.*, 1979] to the IPCC report [IPCC, 2013].

Improving parameterizations of these fast-physics processes is thus essential to reducing uncertainty in climate simulations and to increasing the ability in the projections of future climate. For this purpose, the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) Program established observational sites over an area of a typical GCM grid cell. High-resolution, surface-based measurements are gathered for characterizing a variety of important atmospheric processes [Stokes and Schwartz, 1994; Ackerman and Stokes, 2003]. Along with high-resolution measurements, a hierarchy of models is commonly used, including single-column models (SCMs), cloud-resolving models (CRMs), and large eddy simulations (LES) [e.g., Zhang and Lin, 1997; Randall and Cripe, 1999; Song *et al.*, 2013]. The observing and modeling efforts collectively aim to understand fast processes on a variety of physical scales within a GCM grid cell, and evaluate and improve associated parameterizations. At its outset, the ARM program identified a specific strategy "to develop methods that will allow the output of individual instruments, which measure different parameters, to be combined to infer the time-dependent three-dimensional field of meteorological variables" [Stokes and Schwartz, 1994]. The fields that are inferred can then be used to test GCMs on a variety of scales and, toward this end, data assimilation was explicitly suggested by Stokes and Schwartz [1994].

Data assimilation is a methodology based on optimal estimation theory [Ménard and Daley, 1996; Cohn, 1997; Li and Navon, 2001], which attempts to integrate all available observations into a model to produce analysis fields that can be used to provide model initial conditions to improve forecasts, perform diagnostic analyses,

as well as other applications. The meteorological community has employed data assimilation for more than three decades to provide initial conditions for numerical weather prediction models and to develop reanalysis products for a wide spectrum of applications [Kalnay, 2003].

In this study, we aim to produce fine-resolution, three-dimensional analysis fields for the ARM Southern Great Plains (SGP) site using the Community Gridpoint Statistical Interpolation (GSI) data assimilation system (<http://www.dtcenter.org>). The GSI system is based on a three-dimensional variational data assimilation (3DVAR) algorithm. This system was evolved from its predecessor that is known as the Spectral Statistical Interpolation and was developed in the late 1980s at the National Centers for Environmental Prediction (NCEP) [Parrish and Derber, 1992]. The GSI system has been used operationally at NCEP for about two decades, and it continues to be improved in its performance and capability of assimilating new measurements [Wu et al., 2002; Kleist et al., 2009].

In this application, we implement GSI in a multiscale data assimilation (MS-DA) framework [Li et al., 2012]. The MS-DA algorithm is formulated for fine-resolution models at a resolution down to an order of 1 km [Li et al., 2012; Toth et al., 2013]. Such fine-resolution models encompass a wide range of temporal and spatial scales. In general, data assimilation algorithms attempt to minimize a cost function to obtain a minimum variance, or maximum likelihood estimation, known as the optimal estimation [e.g., Cohn, 1997]. When data assimilation is applied to a fine-resolution model, small-scale structures are subjected to strong filtering effects [Daley, 1991]. A few studies have demonstrated that a set of data assimilation should be applied for a sequence of reduced decorrelation length scales [e.g., Xie et al., 2011; Zhang et al., 2011]. In the MS-DA algorithm, the cost function is decomposed for distinct spatial scales. Here the cost function is decomposed into a large-scale and small-scale component. One advantage of MS-DA is that it solves the data assimilation problem sequentially from large to small scales to reduce the filtering on small scales, thus enabling enhanced constraints on small scales through the assimilation of high-resolution observations.

Another advantage of using the MS-DA algorithm is to exploit existing reanalysis products developed at meteorological centers in addition to assimilating high-resolution measurements. We implement the MS-DA algorithm in the Weather Research and Forecasting (WRF) model at a cloud-resolving resolution of 2 km, which is a much finer resolution than those used in reanalysis products. In the present work, we use the North American Regional Reanalysis (NARR) [Mesinger et al., 2006], which has a horizontal resolution of 32 km. In the MS-DA algorithm, we use the NARR reanalysis as the large-scale component, and the MS-DA focuses on the small-scale component.

This paper describes the implementation of the MS-DA algorithm and presents evaluations of the generated analysis using a variety of cloud and precipitation observations. Section 2 presents the MS-DA algorithm, its implementation, and model configuration. In section 3, we assess the quality of the generated MS-DA analysis using independent observations. Noting that large-scale forcing is needed to drive SCMs, CRMs, and LES, in section 4 we derive large-scale forcing from the generated MS-DA analysis. An operational product of large-scale forcing has been developed and extensively used in the ARM program [Zhang and Lin, 1997; Randall and Cripe, 1999; Ghan et al., 2000; Xie et al., 2004; Fridlind et al., 2012], and the derived large-scale forcing is further evaluated against the operational product [Xie et al., 2010]. Summary and discussion are given in section 5.

Two companion papers (S. Feng et al., Development of fine-resolution analyses and expanded large-scale forcing properties: 2. Scale awareness and application to single-column model experiments, submitted to *Journal of Geophysical Research*, 2014; S. Feng et al., Development of fine-resolution analyses and expanded properties of large-scale forcing: 3. Hydrometeor forcing and application to single-column model experiments, submitted to *Journal of Geophysical Research*, 2014) extend this work by using the generated fine-resolution, three-dimensional fields to explore the contributions of subgrid variability and hydrometeor forcing to the large-scale forcing.

2. Multiscale Data Assimilation

For completeness, we present the basic formulation of MS-DA and then describe its implementation in this application.

2.1. The MS-DA Algorithm

The MS-DA framework is formulated based on the 3DVAR algorithm, which seeks an analysis x^a that minimizes the cost function with respect to the state variable x

$$J(x) = \frac{1}{2}(x - x^b)^T B^{-1}(x - x^b) + \frac{1}{2}(y - Hx)^T R^{-1}(y - Hx) \quad (1)$$

In this cost function, x is the N vector, and y is the M vector consisting of observations. x^b is known as the background, and B is the $N \times N$ background error covariance given by

$$B = \langle e^b (e^b)^T \rangle,$$

where $\langle \rangle$ denotes an ensemble mean over many realizations, and the superscript T stands for transpose. Here $e^b = x^b - x^t$ is the background error vector, where the superscript t indicates the true state. The M vector y consists of observations, and the $M \times M$ matrix R is the observation error covariance associated with the observation vector y . The $M \times N$ matrix H is an observational operator that maps the model state variable to the observation and is assumed to be linear to simplify the discussion here. This analysis is statistically optimal as a minimum error variance estimate [Jazwinski, 1970; Cohn, 1997] or is a maximum likelihood (Bayesian) estimate if both the forecast and observation errors have Gaussian distributions.

Following the notation suggested by Ide et al. [1997], equation (1) can be written in the incremental form

$$J(\delta x) = \frac{1}{2} \delta x^T B^{-1} \delta x + \frac{1}{2} (H \delta x - d)^T R^{-1} (H \delta x - d) \quad (2)$$

where $\delta x = x - x^b$ denotes the increment of the state variable, and $d = y - Hx^b$ is the innovation.

To represent the multiscale nature, δx is further partitioned into two components of spatially distinct scales that yields,

$$\delta x = P_L \delta x_L + P_S \delta x_S = \delta x_L + \delta x_S, \quad (3)$$

where δx_L and δx_S denote the large- and small-scale components of δx , respectively; P_L and P_S are the corresponding linear operators and can be spatial filters or orthogonal decompositions.

Corresponding to equation (3), the background error can be decomposed as

$$e^b = P_L e^b_L + P_S e^b_S = e^b_L + e^b_S, \quad (4)$$

where e_L and e_S are the large- and small-scale components of the background error. Following equation (4), we obtain

$$B = \langle e^b_L (e^b_L)^T \rangle + \langle e^b_S (e^b_S)^T \rangle = B_L + B_S, \quad (5)$$

where B_L and B_S are the error covariances associated with x^b_L and x^b_S . Background error covariances of the additive form as in equation (5) have been used to improve the effectiveness of 3DVAR in assimilating high-resolution observations [Wu et al., 2002]. To obtain equation (5), we have assumed that the large and small-scale background errors are uncorrelated; that is, $\langle e_L (e_S)^T \rangle = 0$.

With the decomposition given in equations (3) and (5), we follow Li et al. (A multiscale data assimilation scheme: Formulation and illustration, submitted to *Monthly Weather Review*, 2014) to decompose equation (2) into two cost functions

$$J_L(\delta x_L) = \frac{1}{2} \delta x_L^T B_L^{-1} \delta x_L + \frac{1}{2} (H \delta x_L - d)^T (R + H B_S H^T)^{-1} (H \delta x_L - d), \quad (6)$$

$$J_S(\delta x_S) = \frac{1}{2} \delta x_S^T B_S^{-1} \delta x_S + \frac{1}{2} (H \delta x_S - d)^T (R + H B_L H^T)^{-1} (H \delta x_S - d), \quad (7)$$

We refer to the 3DVAR that uses the partitioned cost functions equations (6) and (7) as MS-3DVAR, in short "MS-DA."

The background error covariance is characterized by the decorrelation length scale, and this length scale dictates the scales beyond which processes are filtered out [Daley, 1991]. The smaller the decorrelation length scale is, the more effectively high-resolution observations are assimilated and, thus, a small decorrelation length scale in B_S enhances the effectiveness of the assimilation of high-resolution observations.

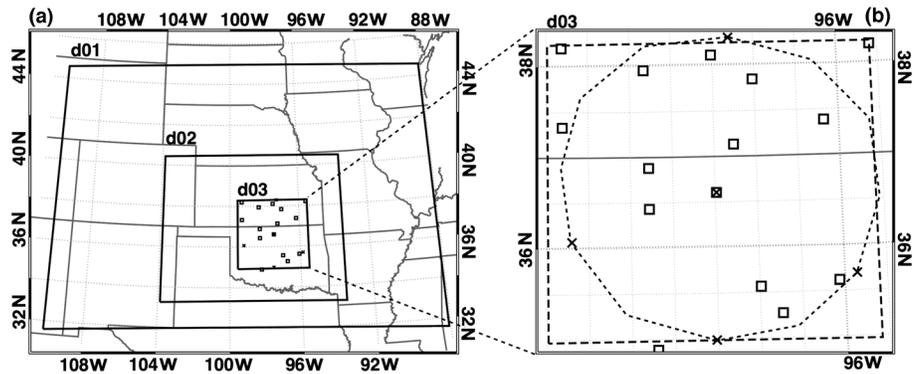


Figure 1. (a) Three nested WRF domains with a resolution of 18 km, 6 km, and 2 km and (b) the analysis domain with the locations of the ARM SONDE (crosses) and SMOS (squares). The long-dashed square box in Figure 1b indicates the area where the large-scale forcing fields are derived. The short-dashed dodecagon in Figure 1b shows the domain that is used for generating the ARM forcing product.

2.2. MS-DA Implementation

As discussed in section 1, one advantage of using the MS-DA algorithm is that it can make use of existing reanalysis products. Current regional reanalysis generally has a resolution on the order of 10 km, which are much coarser than the cloud-resolving resolution that we seek. To proceed, we assume that the regional reanalysis is the large-scale component associated with the cost function equation (6), and the small-scale component analysis is obtained by minimizing the cost function equation (7).

The cost function given in equation (7) has the same form as that given in equation (2) for 3DVAR, but the error covariances are different. In the observational error covariance, an additional term $HB_L H^T$ appears, and this term is known as representativeness error covariance. This correspondence in the form of the cost function allows us to use an existing 3DVAR system for the small-scale data assimilation. We chose here the GSI system, with a modified error covariance.

The GSI system is relatively straightforward to implement in the MS-DA algorithm because of its unique scheme of constructing the background error covariance. It uses a recursive filtering method to construct the background error correlations whereby a decorrelation length scale can be explicitly specified [Wu *et al.*, 2002]. This is necessary for small-scale data assimilation to be performed using equation (7).

The primary motivation for us to use the GSI system is that it is an operational system that has been extensively evaluated on the daily basis. More importantly, it has been developed with an unprecedented capability of assimilating a wide range of observations during the past two decades [Derber and Wu, 1998; Kleist *et al.*, 2009]. It can assimilate basically all types of the conventional observations and a variety of satellite radiances. The assimilation of satellite radiances is particularly desirable for the present work, since it helps constrain radiation balances in the analysis.

Note that there are data assimilation schemes more advanced than 3DVAR, such as four-dimensional variational data assimilation (4DVAR) and 3/4DVAR-based ensemble-variational hybrid data assimilation [Lorenç, 2003; Clayton *et al.*, 2013]. Although 3DVAR is still the dominant scheme for regional models, a GSI-based ensemble-variational hybrid data assimilation has also been developed [Wang *et al.*, 2013]. Once the GSI-based hybrid scheme becomes available, it is straightforward to update the MS-DA methodology presented for use here.

2.3. Model Configuration

Version 3.4 of the WRF model is employed in this study. The model is configured with triple-nested domains that are roughly centered on the ARM SGP central facility (36.6°N, 97.5°W) (Figure 1a). The grid spacing of the nested domains, from the largest area to the smallest, is 18, 6, and 2 km. The three domains all have 45 vertical layers with the top at 100 hPa. The Morrison double-moment microphysics [Morrison and Gettelman, 2008], Yonsei State University planetary boundary layer physics [Noh *et al.*, 2003], and the Noah land surface model are used. The Kain-Fritsch cumulus scheme [Kain, 2004] is applied to the outer and middle domains,

but no cumulus scheme is applied to the inner domain due to its cloud-resolving resolution (2 km). In the discussion that follows, we focus on the innermost domain.

As mentioned previously, NARR [Mesinger *et al.*, 2006] is used as the large-scale data assimilation reanalysis. NARR has a spatial resolution of 32 km and contains temperature, wind, moisture, soil data, and dozens of other parameters, which result from assimilating a large amount of observational data to produce a long-term three-dimensional data set over North America. The observations that are assimilated into NARR include temperatures, winds, and moisture from radiosondes as well as pressure data from surface observations. Also included in the data set are dropsondes, aircraft temperatures and winds, satellite radiances (a measure of heat) from polar orbiting satellites, and cloud drift winds from geostationary satellites.

Since the innermost domain in the WRF configuration has a spatial resolution of 2 km, it is 1 order of magnitude finer than that of NARR. Thus, the small-scale background $x_s^b = 0$, at least for spatial scales smaller than 32 km in the NARR reanalysis. The small-scale component is estimated from the innovation d ; however, the available observations are often inadequate to fully constrain the small-scale component, because the dimension of x_s is generally much larger than the dimension of d . To mitigate this inadequacy, we combine the small-scale data assimilation with downscaling.

Small-scale components can often be reproduced from a prescribed large-scale component through nonlinear dynamical interactions in the model, as demonstrated in downscaling simulations [Castro *et al.*, 2005; Shapiro *et al.*, 2010]. Here the downscaling simulation is achieved by initializing WRF with the NARR data and integrating the WRF model for a 12 h period. After the downscaling integration, the small-scale data assimilation is applied at 6 h intervals during the subsequent 24 h model integration. The four data assimilation analyses generated during the subsequent 24 h are the MS-DA analyses that we aim to produce. For example, the NARR data are used to initialize the WRF model at 12 UTC 12 June 2007, and the model is integrated for 12 h. Small-scale data assimilation is applied at 00 UTC, 06 UTC, 12 UTC, and 18 UTC 13 June 2007. The data assimilation analyses at these four times are the MS-DA analyses. In the following sections, we also use hourly fields. The hourly fields are the model forecasts, filling the gaps between the 6-hourly analyses.

In the experiments presented, the ARM observations assimilated include those from Surface Meteorological Observational Stations (SMOS) and vertical profiles from balloon-borne sounding system (SONDE). The locations of the observation sites are shown in Figure 1b. Along with these ARM observations, measurements operationally assimilated by NCEP are also used, such as conventional observations from radiosondes and radiances from an array of satellites [Kleist *et al.*, 2009].

Given the density of observations, a model resolution of 2 km, and a NARR resolution of 32 km, we specify the decorrelation length scale as 30 km. This scale is given close to the NARR resolution along with an empirical adjustment. Using this small decorrelation length scale, we attempt to limit the filtering on the scales larger than 60 km (twice the decorrelation length scale) and thus assimilate the high-resolution observations as effectively as possible.

3. Evaluation of the MS-DA Analysis

We present results associated with a convective cloud and precipitation event, which occurred from 13 to 15 June 2007 during the Cloud and Land Surface Interaction Campaign at SGP. During this time, a typical mesoscale convective system (MCS) [Houze, 2004] formed, intensified, and decayed. The event thus offers an ideal case for evaluating the representations of convective cloud and precipitation by the MS-DA analysis. The evaluation is conducted primarily by comparing modeled clouds and precipitation against observations. Further, to demonstrate the effectiveness of the MS-DA in reproducing the MCS, we also run control WRF simulations without the MS-DA. The difference between the runs with and without the MS-DA represents the improvement due to MS-DA.

3.1. Reproduction of the Mesoscale Convective System

An MCS moved into the ARM SGP site from the northwest around 21 UTC, 13 June 2007 and intensified. Figure 2 presents a GOES (Geostationary Operational Environmental Satellite) infrared image and a NEXRAD (Next-Generation Radar) reflectivity structure of the MCS around 03 UTC 14 June.

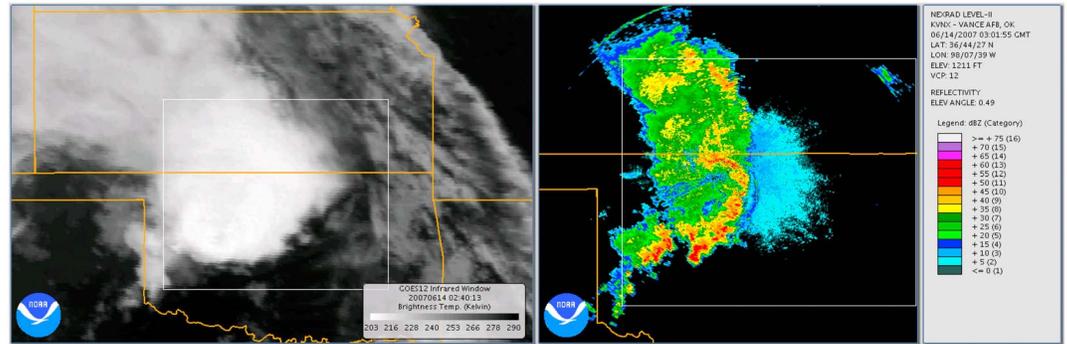


Figure 2. (left) GOES infrared image and (right) NEXRAD reflectivity map. The GOES image is a snapshot at 0240 UTC 14 June. The observed reflectivity map was taken at 0302 UTC 14 June with an elevation angle of 0.49°. The radar is located at 36.74°N, 98.13°W.

The WRF model is initialized at 12 UTC 12 June 2007. Figure 3 displays the simulated maximum reflectivity with and without MS-DA 39 h later at 03 UTC 14 June. Without MS-DA (Figure 3b), the model reflectivity is significantly weaker than observed, and the MCS structure is loosely organized. In contrast, with the MS-DA (Figure 3a), a strong convective echo is realistically reproduced in both its intensity and spatial structure. We conclude that the MS-DA significantly improves the representation of the MCS.

3.2. Hydrometeor Reflectivity

The ARM SGP site is equipped with a set of active remote sensors, such as a millimeter wavelength cloud radar, micropulse lidar, and ceilometer. Combining the measurements from these instruments, the ARM Active Remote Sensing of Clouds (ARSCL) product provides estimates of vertical profiles of hydrometeor reflectivity over SGP [Clothiaux *et al.*, 2000, 2001]. We compare the profiles calculated from the WRF model with those derived from the radar measurements.

In Figure 4a, we see two strong events in the ARSCL radar reflectivity profiles, one from 12 to 15 UTC 13 June and the other from 00 to 06 UTC 14 June. Figure 4b shows that the MS-DA analysis (averaged over 100 km) reproduces the two events fairly well. For the simulation without MS-DA, the reflectivity is significantly underestimated (Figure 4c). The contrast suggests that the MS-DA improves the representation of hydrometeor reflectivity.

3.3. Precipitation

We further evaluate the modeled precipitation using the data from the ABRFC (Arkansas-Red Basin River Forecast Center, available at <http://www.arm.gov/data/vaps/abrfc>). This data consists of 4 km hourly precipitation determined from a combination of WSR-88D NEXRAD radar precipitation estimates and rain

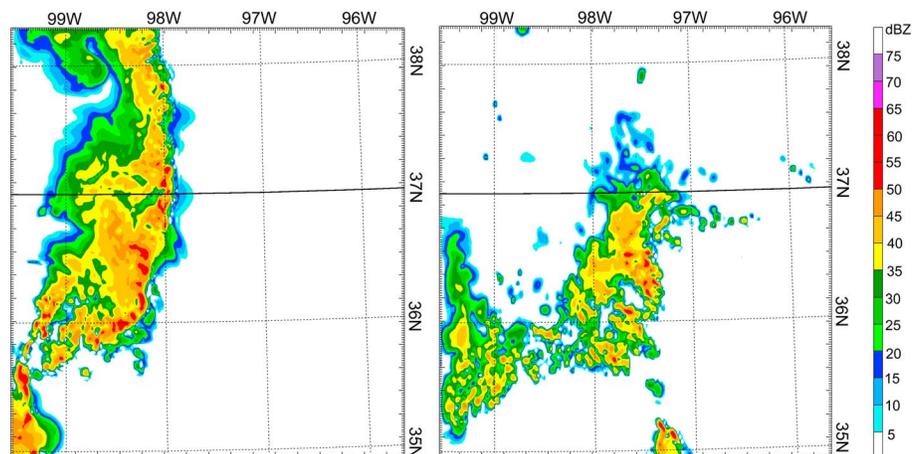


Figure 3. Modeled maximum reflectivity at 03 UTC 14 June 2007. Reflectivity obtained (a) with MS-DA and (b) without MS-DA. The color scale indicates reflectivity in dBZ.

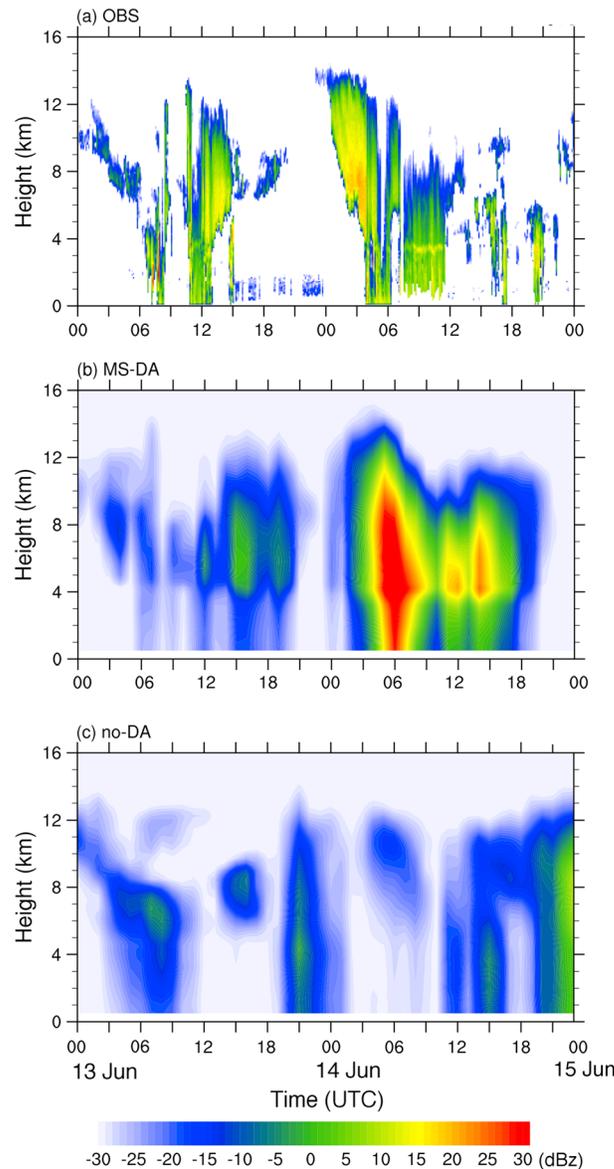


Figure 4. Time evolution of hydrometeor reflectivity vertical profile from (a) ARSCL, (b) simulation with MS-DA, and (c) simulation without MS-DA. The ARSCL data are 10 min averages with a vertical resolution of 45 m, from 105 m to 23,100 m. The hydrometeor reflectivity in Figures 4b and 4c were calculated using the hourly forecasts and averaged over a 100 km × 100 km domain centered on the SGP central facility.

2005]. From the fine-resolution MS-DA analysis, we can derive large-scale forcing fields. We evaluate the MS-DA analysis by comparing the derived large-scale forcing fields to those in the ARM product.

4.1. Formulation of Large-Scale Forcing

To derive large-scale forcing fields from the fine-resolution MS-DA analysis, we follow the formulation by Zhang and Lin [1997]. We write the governing equations of the large-scale atmospheric fields in the form

$$\frac{\partial \bar{V}}{\partial t} + \bar{V} \cdot \nabla \bar{V} + \bar{\omega} \frac{\partial \bar{V}}{\partial p} + \mathbf{f} \mathbf{k} \times \nabla \bar{V} + \nabla \bar{\Phi} = -\nabla \bar{V}' \bar{V}' - \frac{\partial \bar{\omega}' \bar{V}'}{\partial p} \quad (8)$$

$$\frac{\partial \bar{T}}{\partial t} + \bar{V} \cdot \nabla \bar{T} + \bar{\omega} \left(\frac{\partial \bar{T}}{\partial p} - \frac{\bar{\alpha}}{c_p} \right) = \frac{\bar{Q}}{c_p} - \nabla \cdot \bar{V}' \bar{T}' - \left(\frac{\partial \bar{\omega}' \bar{T}'}{\partial p} - \frac{\bar{\omega}' \bar{\alpha}'}{c_p} \right), \quad (9)$$

gauge reports. Figure 5 displays a time series of precipitation rate from the ABRFC data and from the WRF simulation with and without MS-DA. The model precipitation rates are averages over the innermost domain shown in Figure 1b. Two precipitation events are observed during the period, centered at 15 UTC 13 June and 06 UTC 14 June. For the precipitation simulation without MS-DA, the peaks during the major precipitation events are lower than observed and the maximum rate lags the observation by about 9 h. In contrast, the modeled precipitation with MS-DA occurs at nearly the same time as in the observations for both events. While slightly underestimating the peak during the first precipitation event (18 UTC 13 June), MS-DA reproduces the amplitude of the second event well (00–12 UTC 14 June).

Figure 6 illustrates the spatial distribution of hourly precipitation at 08 UTC 13 June and 06 UTC, and 18 UTC 14 June. These three snapshots are selected to depict precipitation during the development of the first precipitation event, at the peak of the second event, and after the second event, respectively. The results suggest that the MS-DA significantly improves the model precipitation.

4. Comparison With Large-Scale Forcing Fields

For the ARM SGP site, the large-scale forcing product has been produced using a constrained objective variational analysis [Zhang and Lin, 1997]. This forcing is a primary data product for the ARM program and has been carefully evaluated and extensively used [e.g., Ghan et al., 2000; Xie et al., 2003; Xie et al.,

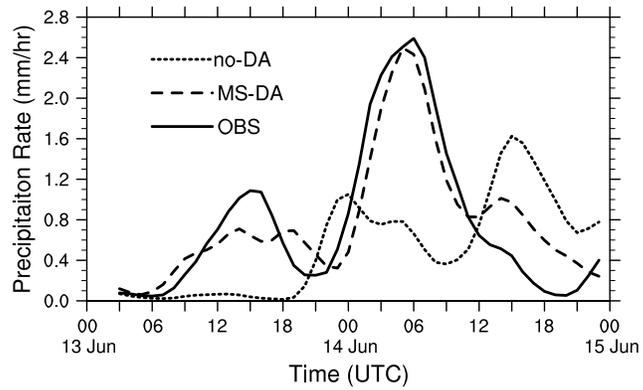


Figure 5. Time series of domain-averaged precipitation rate (mm/h). Solid line denotes the ABRFC observation; dashed denotes the simulation with MS-DA; and the dotted line denotes the simulation without data assimilation.

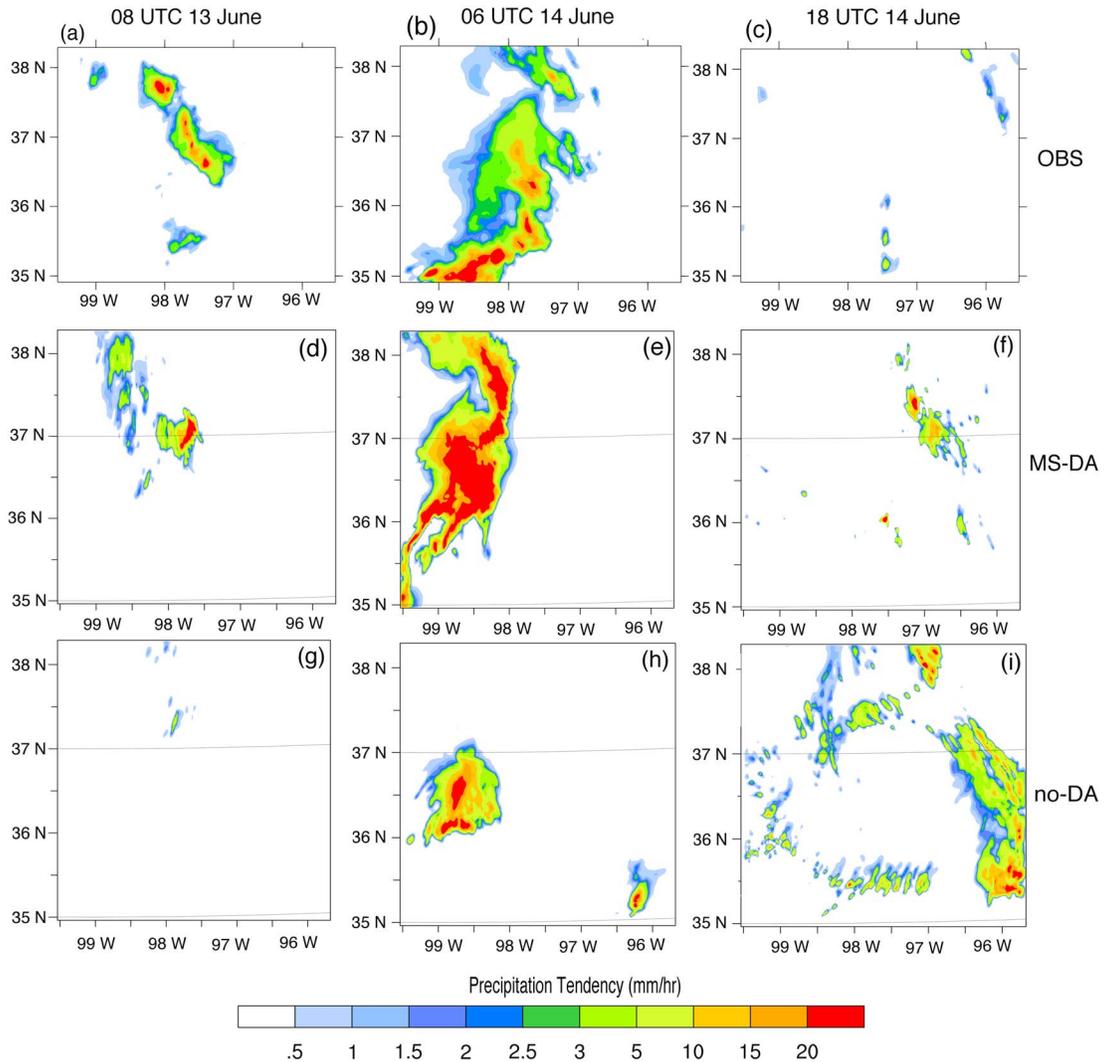


Figure 6. Horizontal distribution of precipitation rate (mm/h) at 08 UTC 13 June and at 06 UTC and 18 UTC June 14 from (a–c) ABRFC observation, (d–f) MS-DA, and (g–i) no DA experiments.

$$\frac{\partial \bar{q}}{\partial t} + \bar{\nabla} \cdot \bar{\nabla} \bar{q} + \bar{\omega} \frac{\partial \bar{q}}{\partial p} = \bar{S} - \overline{\nabla \cdot \mathbf{V}' \bar{q}} - \frac{\partial \overline{\omega' \bar{q}}}{\partial p}, \quad (10)$$

$$\bar{\nabla} \bar{\omega} + \frac{\partial \bar{\omega}}{\partial p} = 0, \quad (11)$$

where the overbar denotes the horizontal average over a specified domain that represents a GCM grid box, and the prime deviation from the domain average. The variables are defined as follows: \mathbf{V} , horizontal wind; T , air temperature; q , mixing ratio of water vapor; ω , vertical p velocity; p , pressure; α , specific volume of the air; c_p , the specific heat at constant volume; ϕ , geopotential; f , Coriolis parameter; Q , the heating rate; S , the source of water vapor; and \mathbf{k} , the unit vector in the direction pointing upward. ∇ is the horizontal del operator.

Following equations (8)–(11), the large-scale forcing fields are defined as

$$\left(\frac{\partial \bar{\mathbf{V}}}{\partial t} \right)_{LS} = -\bar{\mathbf{V}} \cdot \bar{\nabla} \bar{\mathbf{V}} - \bar{\omega} \frac{\partial \bar{\mathbf{V}}}{\partial p} - f \mathbf{k} \times \bar{\nabla} \bar{\mathbf{V}} - \bar{\nabla} \bar{\Phi} \quad (12)$$

$$\left(\frac{\partial \bar{T}}{\partial t} \right)_{LS} = -\bar{\mathbf{V}} \cdot \bar{\nabla} \bar{T} - \bar{\omega} \left(\frac{\partial \bar{T}}{\partial p} - \frac{\bar{\alpha}}{c_p} \right), \quad (13)$$

$$\left(\frac{\partial \bar{q}}{\partial t} \right)_{LS} = -\bar{\mathbf{V}} \cdot \bar{\nabla} \bar{q} - \bar{\omega} \frac{\partial \bar{q}}{\partial p}, \quad (14)$$

$$\bar{\omega} = \int_{p_0}^p \bar{\nabla} \cdot \bar{\mathbf{V}} dp, \quad (15)$$

where p_0 is the pressure at the surface.

Given the large-scale forcing fields, equations (8)–(10) can be integrated in time within a single column in isolation from the model. They thus consist of the basic equations of a SCM. We can also see that equations (9) and (10) are independent of equation (8). This implies that an SCM lacks dynamical feedbacks that occur within complete three-dimensional atmospheric models. Practically speaking, equations (9) and (10) can be integrated separately. In the following discussion, we are concerned only with equations (9) and (10).

We note that the terms associated with horizontal velocities, \mathbf{V}' , arise from subgrid processes that are not resolved in GCMs. They are partially parameterized as hyperdiffusion in most GCMs [e.g., Palmer, 2001]. In large-scale forcing fields, they are generally included [Bechtold *et al.*, 2000] or modeled using a nudging term [Randall and Cripe, 1999]. The contribution of this subgrid variability to large-scale forcing is not fully understood and will be addressed in the companion paper (S. Feng *et al.*, Development of fine-resolution analyses and expanded large-scale forcing properties: 2. Scale awareness and application to single-column model experiments, submitted manuscript, 2014). In the following sections, we calculate the large-scale forcing fields for temperature, water vapor, and vertical velocity following equations (13)–(15).

4.2. Derived Large-Scale Forcing

Figure 7 presents the derived large-scale forcing fields along with those from the ARM forcing product. The observed precipitation rate is overlaid on the vertical velocity plots. Overall, the time evolution of the derived large-scale forcing agrees well with that from the ARM forcing product. In Figures 7a and 7b, we see an intense upward motion event 00~12 UTC 14 June. It is associated with intense cooling in the middle and upper troposphere (Figures 7c and 7d) and with high moisture content in the lower troposphere (Figures 7e and 7f). Strong precipitation occurs during the strong upward motion event (Figures 7a and 7b), which indicates a large-scale balance between atmospheric motions and precipitation.

4.3. Single-Column Model Experiments

To further evaluate the fine-resolution MS-DA analysis, we examine SCM simulations driven by the derived large-scale forcing. The SCM experiments are conducted using the single-column version of the National Center for Atmospheric Research (NCAR) Community Atmospheric Model version 5 (CAM5), hereafter referred to as SCAM5. The SCAM5 model contains the vertical advection scheme and all of the physics routines used in CAM5 [Neale *et al.*, 2012], including the cloud microphysics and cloud macrophysics

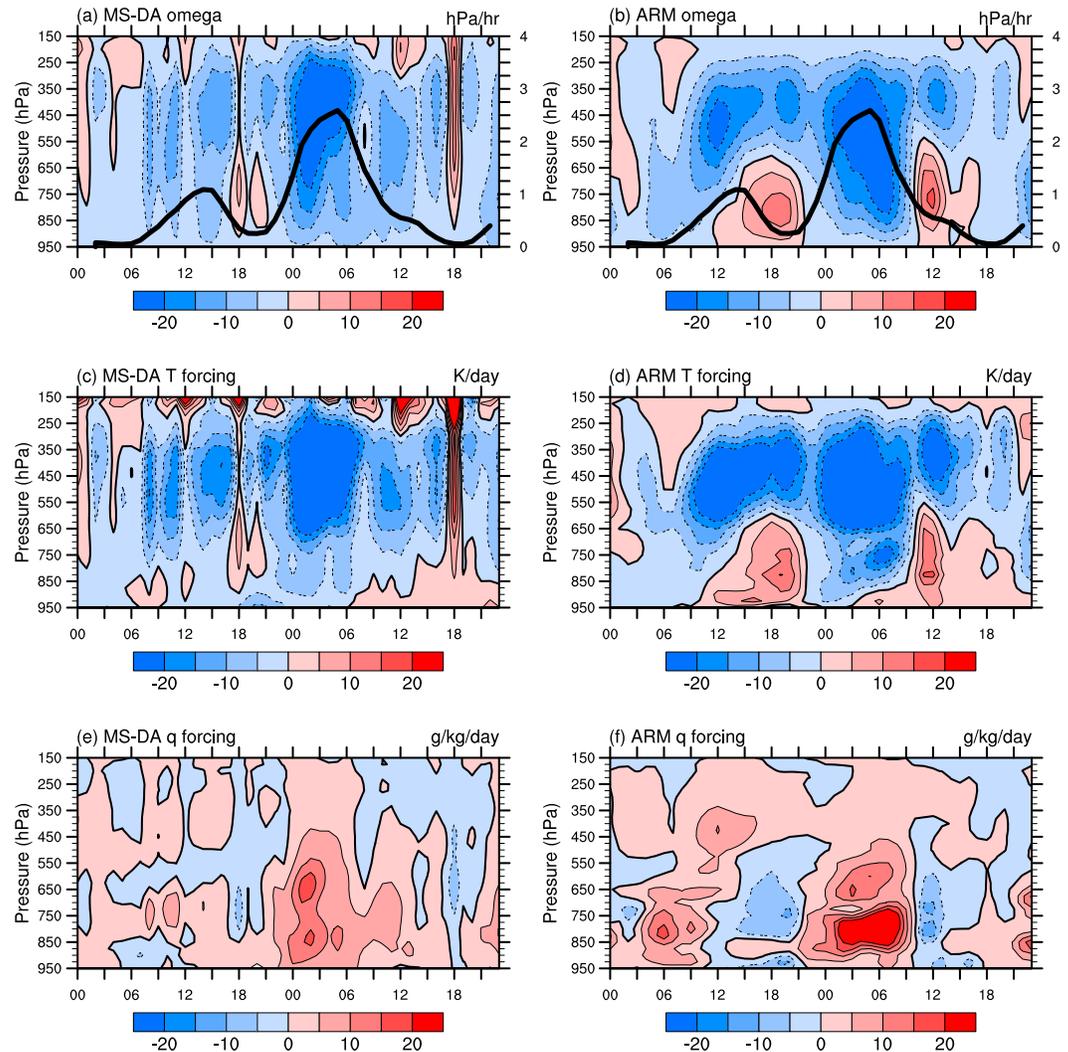


Figure 7. Time-pressure distribution of (a and b) the large-scale vertical velocity $\bar{\omega}$, (c and d) the large-scale T forcing, namely, $\left(\frac{\partial T}{\partial t}\right)_{LS}$, and (e and f) the large-scale q forcing, namely, $\left(\frac{\partial q}{\partial t}\right)_{LS}$. The thick solid lines in Figures 7a and 7b are the observed surface precipitation rates (mm/h).

schemes. A detailed description of these schemes can be found in Gettelman *et al.* [2008], Morrison and Gettelman [2008], and Gettelman *et al.* [2010]. Note that no relaxation/nudging is applied in the experiments presented here.

Figure 8 presents the SCAM precipitation rates from simulations driven by the MS-DA-derived large-scale forcing and the ARM forcing product. Also shown are the observed and MS-DA-simulated precipitation rates given in Figure 5. The simulations by the two forcings produce precipitation patterns that are very similar, confirming the consistency between the derived large-scale forcing and the ARM forcing product.

Comparing the simulations to the observations, both simulations capture the major precipitation events and reproduce the overall time evolution, but there are two noticeable limitations. First, the simulated precipitation lags the observations for both forcings. For the MS-DA-derived forcing, the lag is about 3 h; for the ARM forcing, the lag is somewhat longer, about 3–6 h. Second, the simulations overpredict the peak intensity of the precipitation. We note that these two limitations are not necessarily attributable to deficiencies in the cloud and precipitation physics parameterization in the CAM5, since they could arise from the uncertainties in the large-scale forcing fields. In fact, the overprediction of precipitation may be attributed partially to an underestimation of the forcing component from

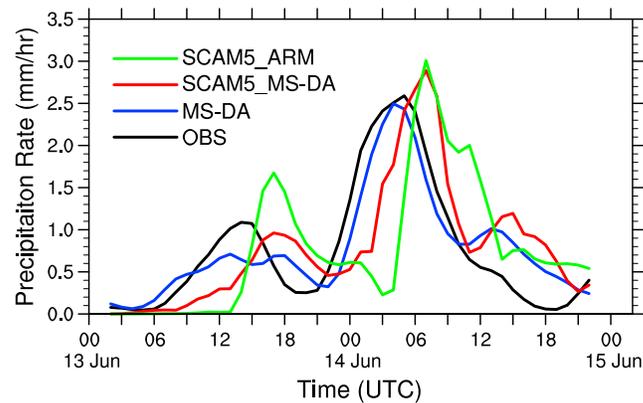


Figure 8. Time series of hourly precipitation rates. The black line denotes the ABRFC observations, and blue line denotes the MS-DA simulation. Both are averaged over the innermost domain (d03) in Figure 1. Red and green lines denote the SCAM5-simulated precipitation driven by the derived large-scale forcing and by the ARM forcing, respectively. The 20 min SCAM5 output is averaged to obtain hourly precipitation rates.

fine-resolution analyses and expanded properties of large-scale forcing: 3. Hydrometeor forcing and application to single-column model experiments, submitted manuscript, 2014).

5. Discussion and Summary

Data assimilation was recognized as a basic strategy in the ARM program at its outset nearly two decades ago. In the ARM program, the generation of the three-dimensional fields from the observations remains challenging even for SGP, the most instrumented ARM site. One reason is that the observations acquired by the ARM program are insufficient to fully constrain the three-dimensional fields down to a cloud-resolving scale. This motivated the development of the constrained objective variational analysis [Zhang and Lin, 1997], which generates large-scale forcing fields rather than three-dimensional fields.

Data assimilation has progressed greatly over the past two decades and has continued to enhance capabilities to assimilate additional observations, particularly satellite radiances. At the same time, regional reanalyses have become available and regional modeling, represented by the community WRF model, has rapidly advanced. By using the MS-DA algorithm, we attempt to capitalize on this progress and examine whether we can produce fine-resolution three-dimensional fields at cloud-resolving scales that are useful for practical applications within the ARM program.

We have assimilated ARM observations along with measurements from other observing networks into a WRF model at a cloud-resolving resolution of 2 km over the ARM SGP site. The GSI data assimilation system is implemented in a MS-DA framework, and it has been applied to a set of cases for a variety of cloud and precipitation regimes. The case presented here is for a challenging mesoscale convective system. The results obtained are encouraging—the performance assessments show that MS-DA significantly improves the representation of the intensity and structure of precipitation and clouds associated with the MCS.

For further evaluation of the MS-DA analysis, we derived large-scale forcing fields from high-resolution, three-dimensional fields and compared it with the ARM large-scale forcing product [Xie *et al.*, 2004]. The comparison shows that the derived large-scale forcing has an overall accuracy comparable to the ARM forcing. The robustness of this result is confirmed using a set of SCM simulations. The comparable accuracy between the derived large-scale forcing and the ARM forcing product motivates us to explore in detail the properties of the large-scale forcing, such as subgrid variability and hydrometeor forcing.

The spatial resolution of climate models has been rapidly increasing in recent years. Some climate models now have a resolution on the order of 10 km. As a consequence, scale aware parameterizations have been under intensive development [Arakawa and Jung, 2011; Grell and Freitas, 2013]. In order to use SCMs to evaluate such parameterizations with the rapidly increasing resolution of climate models, a corresponding scale aware forcing must be developed. Based on the fine-resolution MS-DA analysis, we can derive forcing

subgrid-scale horizontal advection that appears in equations (9) and (10) (S. Feng *et al.*, Development of fine-resolution analyses and expanded large-scale forcing properties: 2. Scale awareness and application to single-column model experiments, submitted manuscript, 2014). These subgrid-scale horizontal advection components can significantly reduce the precipitation rate and improve the timing of the precipitation occurrence in this case. Furthermore, an even more important issue is that there is no hydrometeor forcing included in the large-scale forcing fields, which we have found can also significantly affect the precipitation rate in this case (S. Feng *et al.*, Development of

on explicitly specified scales and address the impact of grid size in the SCM simulation. Another important issue is the impact of subgrid-scale horizontal advection variability on the large-scale forcing. Leveraging the fine-resolution three-dimensional fields from the MS-DA analysis, this issue can be systematically addressed. The results will be presented in S. Feng et al. (Development of fine-resolution analyses and expanded large-scale forcing properties: 2. Scale awareness and application to single-column model experiments, submitted manuscript, 2014).

Evidence shows that hydrometeor advection can significantly affect cloud water content, specific humidity, temperature, and other fields in SCM simulations [Petch and Dudhia, 1998]. In Petch and Dudhia [1998], the hydrometeor forcing was derived from regional mesoscale model simulations without data assimilation, but their results pointed to limitations in the representation of clouds and precipitation. Encouraged by the capability of MS-DA to improve the representation of clouds and precipitation, we will derive hydrometeor forcing fields and provide a systematic assessment of the impact of hydrometeor forcing on SCM simulations in the companion paper (S. Feng et al., Development of fine-resolution analyses and expanded properties of large-scale forcing: 3. Hydrometeor forcing and application to single-column model experiments, submitted manuscript, 2014).

Acknowledgments

The research described in this publication was supported by the U.S. Department of Energy Earth System Modeling (ESM) program via the FAST-physics System Testbed and Research (FASTER) project www.bnl.gov/faster. The research was carried out, in part, at Jet Propulsion Laboratory (JPL) California Institute of Technology, under a contract with the National Aeronautics and Space Administration (NASA). The authors thank the ARM program for providing the SGP observations. The authors are grateful to Ann Fridlind (NASA Goddard Institute for Space Studies) for numerous stimulating discussions, insightful suggestions, and strong support. The authors thank the anonymous reviewers for comments that were very helpful in improving the manuscript.

References

- Ackerman, T. P., and G. M. Stokes (2003), The atmospheric radiation measurement program, *Phys. Today*, *56*(1), 38–44.
- Arakawa, A., and J. H. Jung (2011), Multiscale modeling of the moist-convective atmosphere—A review, *Atmos. Res.*, *102*(3), 263–285.
- Bechtold, P., et al. (2000), A GCS model intercomparison for a tropical squall line observed during toga-coare. II: Intercomparison of single-column models and a cloud-resolving model, *Q. J. R. Meteorol. Soc.*, *126*(564), 865–888.
- Castro, C. L., R. A. Pielke, and G. Leoncini (2005), Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS), *J. Geophys. Res.*, *110*, D05108, doi:10.1029/2004JD004721.
- Charney, J. G., A. Arakawa, J. D. Baker, B. Bolin, R. E. Dickinson, R. M. Goody, C. E. Leith, H. M. Stommel, and C. I. Wunsch (1979), *Carbon Dioxide and Climate: A Scientific Assessment*, Natl. Acad. Press, Washington, D. C.
- Clayton, A. M., A. C. Lorenc, and D. M. Barker (2013), Operational implementation of a hybrid ensemble/4D-VAR global data assimilation system at the Met Office, *Q. J. R. Meteorol. Soc.*, *139*(675), 1445–1461.
- 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*(5), 645–665.
- Clothiaux, E. E., et al. (2001), The ARM Millimeter Wave Cloud Radars (MMCRs) and the Active Remote Sensing of Clouds (ARSCL) Value Added Product (VAP), *DOE Tech. Memo. ARM VAP-002.1*, U.S. Dep. of Energy, Washington, D. C.
- Cohn, S. E. (1997), An introduction to estimation theory, *J. Meteorol. Soc. Jpn.*, *75*(1B), 257–288.
- Daley, R. (1991), *Atmospheric Data Analysis*, 457 pp., Cambridge Univ. Press, Cambridge, U. K.
- Derber, J. C., and W.-S. Wu (1998), The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system, *Mon. Weather Rev.*, *126*(8), 2287–2299.
- Fridlind, A. M., et al. (2012), A comparison of TWP-ICE observational data with cloud-resolving model results, *J. Geophys. Res.*, *117*, D05204, doi:10.1029/2011JD016595.
- 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*(15), 3660–3679.
- Gettelman, A., X. Liu, S. J. Ghan, H. Morrison, S. Park, A. J. Conley, S. A. Klein, J. Boyle, D. L. Mitchell, and J. L. F. Li (2010), Global simulations of ice nucleation and ice supersaturation with an improved cloud scheme in the Community Atmosphere Model, *J. Geophys. Res.*, *115*, D18216, doi:10.1029/2009JD013797.
- Ghan, S., et al. (2000), A comparison of single column model simulations of summertime midlatitude continental convection, *J. Geophys. Res.*, *105*(D2), 2091–2124, doi:10.1029/1999JD900971.
- Grell, G. A., and S. R. Freitas (2013), A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, *Atmos. Chem. Phys. Discuss.*, *13*(9), 23,845–23,893.
- Houze, R. A. (2004), Mesoscale convective systems, *Rev. Geophys.*, *42*, RG4003, doi:10.1029/2004RG000150.
- Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc (1997), Unified notation for data assimilation: Operational, sequential and variational, *J. Meteorol. Soc. Jpn.*, *75*(1B), 181–189.
- IPCC (2013), Climate change 2013: The physical science basis, in *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., 1535 pp., Cambridge Univ. Press, Cambridge, U. K., and New York.
- Jazwinski, A. H. (1970), *Stochastic Processes and Filtering Theory*, 376 pp., Academic Press, New York.
- Kain, J. S. (2004), The Kain–Fritsch convective parameterization: An update, *J. Appl. Meteorol.*, *43*(1), 170–181.
- Kalnay, E. (2003), *Atmospheric Modeling, Data Assimilation and Predictability*, 341 pp., Cambridge Univ. Press, Cambridge, U. K.
- Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord (2009), Introduction of the GSI into the NCEP global data assimilation system, *Weather Forecasting*, *24*(6), 1691–1705.
- Li, Z., and I. M. Navon (2001), Optimality of variational data assimilation and its relationship with the Kalman filter and smoother, *Q. J. R. Meteorol. Soc.*, *127*(572), 661–683.
- Li, Z., Y. Chao, J. D. Farrara, and J. C. McWilliams (2012), Impacts of distinct observations during the 2009 Prince William Sound field experiment: A data assimilation study, *Cont. Shelf Res.*, doi:10.1016/j.csr.2012.06.018.
- Lorenc, A. C. (2003), The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var, *Q. J. R. Meteorol. Soc.*, *129*(595), 3183–3203.
- Ménard, R., and R. Daley (1996), The application of Kalman smoother theory to the estimation of 4DVAR error statistics, *Tellus A*, *48*(2), 221–237.
- Mesinger, F., et al. (2006), North American regional reanalysis, *Bull. Am. Meteorol. Soc.*, *87*(3), 343–360.
- 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*(15), 3642–3659.

- Neale, R. B., et al. (2012), Description of the NCAR Community Atmosphere Model (CAM 5.0), *Tech. Note NCAR/TN-486 + STR*.
- Noh, Y., W. G. Cheon, S. Y. Hong, and S. Raasch (2003), Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data, *Boundary Layer Meteorol.*, *107*(2), 401–427.
- Palmer, T. N. (2001), A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parameterization in weather and climate prediction models, *Q. J. R. Meteorol. Soc.*, *127*(572), 279–304.
- Parrish, D. F., and J. C. Derber (1992), The national meteorological center's spectral statistical interpolation analysis system, *Mon. Weather Rev.*, *120*, 1747–1763.
- Petch, J. C., and J. Dudhia (1998), The importance of the horizontal advection of hydrometeors in a single-column model, *J. Clim.*, *11*(9), 2437–2452.
- Randall, D., and D. Cripe (1999), Alternative methods for specification of observed forcing in single-column models and cloud system models, *J. Geophys. Res.*, *104*(D20), 24,527–24,545, doi:10.1029/1999JD900765.
- Shapiro, M., et al. (2010), An earth-system prediction initiative for the twenty-first century, *Bull. Am. Meteorol. Soc.*, *91*(10), 1377–1388.
- Song, H., W. Lin, Y. Lin, A. B. Wolf, R. Neggers, L. J. Donner, A. D. Del Genio, and Y. Liu (2013), Evaluation of precipitation simulated by seven SCMs against the ARM observations at the SGP site, *J. Clim.*, *26*(15), 5467–5492.
- Stokes, G. M., and S. E. Schwartz (1994), The atmospheric radiation measurement (ARM) program: Programmatic background and design of the cloud and radiation test bed, *Bull. Am. Meteorol. Soc.*, *75*(7), 1201–1221.
- Toth, Z., M. Tew, D. Birkenheuer, S. Albers, Y. Xie, and B. Motta (2013), Multiscale data assimilation and forecasting, *Bull. Am. Meteorol. Soc.*, *95*(2), ES30–ES33.
- Wang, X., D. Parrish, D. Kleist, and J. Whitaker (2013), GSI 3DVar-based ensemble-variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments, *Mon. Weather Rev.*, *141*(11), 4098–4117.
- Wu, W.-S., R. J. Purser, and D. F. Parrish (2002), Three-dimensional variational analysis with spatially inhomogeneous covariances, *Mon. Weather Rev.*, *130*(12), 2905–2916.
- Xie, S., R. T. Cederwall, M. Zhang, and J. J. Yio (2003), Comparison of SCM and CSRMM forcing data derived from the ECMWF model and from objective analysis at the ARM SGP site, *J. Geophys. Res.*, *108*(D16), 4499, doi:10.1029/2003JD003541.
- Xie, S., R. T. Cederwall, and M. Zhang (2004), Developing long-term single-column model/cloud system-resolving model forcing data using numerical weather prediction products constrained by surface and top of the atmosphere observations, *J. Geophys. Res.*, *109*, D01104, doi:10.1029/2003JD004045.
- Xie, S., et al. (2005), Simulations of midlatitude frontal clouds by single-column and cloud-resolving models during the Atmospheric Radiation Measurement March 2000 cloud intensive operational period, *J. Geophys. Res.*, *110*, D15503, doi:10.1029/2004JD005119.
- Xie, S., et al. (2010), Cloud and more: ARM climate modeling best estimate data, *Bull. Am. Meteorol. Soc.*, *91*(1), 13–20.
- Xie, Y., S. Koch, J. McGinley, S. Albers, P. E. Bieringer, M. Wolfson, and M. Chan (2011), A space-time multiscale analysis system: A sequential variational analysis approach, *Mon. Weather Rev.*, *139*(4), 1224–1240.
- Zhang, F., Y. Weng, J. F. Gamache, and F. D. Marks (2011), Performance of convection-permitting hurricane initialization and prediction during 2008–2010 with ensemble data assimilation of inner-core airborne Doppler radar observations, *Geophys. Res. Lett.*, *38*, L15810, doi:10.1029/2011GL048469.
- Zhang, M. H., and J. L. Lin (1997), Constrained variational analysis of sounding data based on column-integrated budgets of mass, heat, moisture, and momentum: Approach and application to ARM measurements, *J. Atmos. Sci.*, *54*(11), 1503–1524.