Stony Brook University

OFFICIAL COPY

The official electronic file of this thesis or dissertation is maintained by the University Libraries on behalf of The Graduate School at Stony Brook University.

© All Rights Reserved by Author.

An Ensemble Three-Dimensional Constrained Variational Analysis Method to Derive Large-Scale Forcing Data for Single-Column Models

A Dissertation Presented

by

Shuaiqi Tang

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Marine and Atmospheric Sciences

Stony Brook University

May 2015

Stony Brook University

The Graduate School

Shuaiqi Tang

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Minghua Zhang – Dissertation Advisor
Professor & Dean, SOMAS

Brian A. Colle – Chairperson of Defense
Professor & Director, SOMAS

Edmund K.M. Chang – Internal Committee Member
Professor, SOMAS

Marat Khairoutdinov – Internal Committee Member
Associate Professor, SOMAS

Hyemi Kim – Internal Committee Member
Assistant Professor, SOMAS

Shaocheng Xie – External Committee Member
Research Scientist, Atmospheric, Earth and Energy Department
Lawrence Livermore National Laboratory

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

An Ensemble Three-Dimensional Constrained Variational Analysis (3DCVA) Method to Derive Large-Scale Forcing Data for Single-Column Models

by

Shuaiqi Tang

Doctor of Philosophy

in

Marine and Atmospheric Sciences

Stony Brook University

2015

Atmospheric vertical velocities and advective tendencies are essential as large-scale forcing data to drive single-column models (SCM), cloud-resolving models (CRM) and large-eddy simulations (LES). They cannot be directly measured or easily calculated with great accuracy from field measurements. In the Atmospheric Radiation Measurement (ARM) program, a constrained variational algorithm (1DCVA) has been used to derive large-scale forcing data over a sounding network domain with the aid of flux measurements at the surface and top of the atmosphere (TOA). We extend the 1DCVA algorithm into three dimensions (3DCVA) along with other improvements to calculate gridded large-scale forcing data. We also introduce an ensemble framework using different background data, error covariance matrices and constraint variables to quantify the uncertainties of the large-scale forcing data. The results of sensitivity study

show that the derived forcing data and SCM simulated clouds are more sensitive to the background data than to the error covariance matrices and constraint variables, while horizontal moisture advection has relatively large sensitivities to the precipitation, the dominate constraint variable. Using a mid-latitude cyclone case study in March 3rd, 2000 at the ARM Southern Great Plains (SGP) site, we investigate the spatial distribution of diabatic heating sources (Q1) and moisture sinks (Q2), and show that they are consistent with the satellite clouds and intuitive structure of the mid-latitude cyclone. We also evaluate the Q1 and Q2 in analysis/reanalysis, finding that the regional analysis/reanalysis all tend to underestimate the sub-grid scale upward transport of moist static energy in the lower troposphere. With the uncertainties from large-scale forcing data and observation specified, we compare SCM results and observations and find that models have large biases on cloud properties which could not be fully explained by the uncertainty from the large-scale forcing data and observation. The analysis of GCM and satellite data further verifies the model biases globally and climatologically.

Contents

Chapter 1. Introduction 1

Chapter 2. The Method of Three-Dimensional Constrained Variational Analysis 8

2.1 Model of the 3DCVA 8

2.2 Numerical Algorithms 11

Chapter 3. Ensemble 3DCVA and Sensitivity Study 19

3.1 Ensemble 3DCVA 19

3.2 Sensitivity of Large-Scale Forcing Data 26

Chapter 4. A Case Study of Mid-Latitude Cyclone 29

4.1 Synoptic Condition 29

4.2 3D Structure of Q1 and Q2 30

4.3 Sub-grid Transport of Moist Static Energy 32

Chapter 5. Single-Column Model Simulations of Clouds 34

5.1 SCM results of cloud and sensitivities 34

5.2 SCM Model Bias and Source Diagnostics 36

5.3 GCM Verifications 38

Chapter 6. Conclusions and Future Work 41

6.1 Conclusions 41

6.2 Future Work 43

References 46

Appendix 51

Appendix 1: numerical calculation of 1st and 2nd derivatives 51

List of Figures/Tables/Illustrations

Figures

Figure 1.1: Working strategy illustration of GEWEX (Global Energy and Water Cycle Experiment) Cloud System Study (GCSS) presented by A. Pier Siebesma in 2008. Large-scale forcing data derived from field campaigns and/or observational sites are needed to drive LES, CRM and SCM which are used to develop and test the physical parameterizations in GCM and NWP……...59

Figure 2.1: schematic figure of the 3DCVA. It shows the inputs of background data (3D state variables), surface and TOA constraint variables, and the output of 3D large-scale forcing data……60

Figure 2.2: the adjusted C-grid used in 3DCVA. *u* is at the center of east/west grid faces, *v* is at the center of north/south grid faces, *s* and *q* are at the center of all four grid faces and all the constraint variables (precipitation, radiation, surface fluxes) are at the grid center and represent grid average……..61

Figure 2.3: Illustration of horizontal and vertical correlations. *i, j* and *k* represent the horizontal and vertical grid/level indices. Bold *i, j* show the index of the grids while the thin *i, j* show the index of the boundary variables, so the grid (*i, j*) has 4 boundary variables at *i, i+1, j,* and *j+1*. *KB* is the level at the surface and *KT* is the level at the TOA. Blue dots are the variables of the background data, and the red arrows show the correlation of the center variable to its surrounding variables…….62

Figure 3.1: SGP domain (34.75-38.75 °N, 95.25-99.75 °W) averaged profile of (a) precipitation, (b) temperature, (c) water vapor mixing ratio, (d) vertical velocity, (e – h) horizontal and vertical advections of dry static energy *s* and moisture *q*. ……65

Figure 3.2: (left) ensemble mean SGP domain averaged horizontal wind, moisture and dry static energy (*u, v, q, s*) and (right) ensemble mean SGP domain mean adjustments of *u, v, q, s*……..…66

Figure 3.3: the standard deviations of large-scale forcing data (from top to bottom: T, *q*, ω, horizontal and vertical advections of *s* and *q*) due to different background data (left), error covariance matrices (middle) and constraint variables (right, only precipitation)…….67

Figure 4.1: (a). GOES-8 satellite cloud image at 23:45Z March 2nd 2000. The red box shows the SGP domain applying 3DCVA. (b) 3-hourly precipitation rate centered at 00Z

Mar 3rd averaged into 0.5 °×0.5 °horizontal resolution from the Arkansas-Red Basin River Forecast Center (ABRFC). The bold black line shows the position of cross-front section in later figures……68

Figure 4.2: 850hPa Geopotential height (gpm, black lines), temperature (K, red lines) and specific humidity (g/kg, shaded) from (a) RUC, (b) ERA-Interim, (c) JRA 55, (d) NARR, (e) CFSR, (f) MERRA at 00Z Mar 3rd 2000. The red box indicates the SGP domain……69

Figure 4.3: Same as Figure 4.2 except that the shading is for vertical velocity and the level is 500hPa…….……..70

Figure 4.4: The vertically integrated Q1 from (a) ensemble 3DCVA, (b) ensemble background data, (c) RUC, (d) ERA Interim, (e) JRA 55, (f) NARR, (g) CFSR and (h) MERRA at 00Z Mar 3rd 2000. The unit is mm/day. The bold black lines in (a) and (b) shows the position of cross-front section in later figures…….……....71

Figure 4.5: The same as figure 4.4 but for vertical integrated Q2 (mm/day)…….……..72

Figure 4.6: cross section (indicated by the bold black line in figure 4.1(b)) of Q1 (color, K/hour) and circulation along the cross section (streamlines) from (a) ensemble 3DCVA, (b) ensemble background data, (c) RUC, (d) ERA Interim, (e) JRA 55, (f) NARR, (g) CFSR and (h) MERRA. The time is 00Z Mar 3rd 2000. X axis shows the longitude along the cross section……..…….....73

Figure 4.7: cross section (indicated by the bold black line in figure 4.1(b)) of Q2 (color, K/hour) and relative humidity (contour, %) (a) ensemble 3DCVA, (b) ensemble background data, (c) RUC, (d) ERA Interim, (e) JRA 55, (f) NARR, (g) CFSR and (h) MERRA. The time is 00Z Mar 3rd 2000. X axis shows the longitude along the cross section……...……74

Figure 4.8: (top) Q1−Q2 profiles (K/hour) as a function of VIQ1 and (bottom) PDFs of VIQ1 (in fraction) from (a) ensemble 3DCVA, (b) ensemble background data, (c) RUC, (d) ERA Interim, (e) JRA 55, (f) NARR, (g) CFSR and (h) MERRA for all grids in SGP domain during 18Z March 1st to 24Z March 3rd. …….75

Figure 5.1: the ensemble average of simulated cloud properties in SCAM4 (left), SCAM5 (middle) and cloud retrievals from MICROBASE (right). From top to bottom: cloud fraction, LWC and IWC…….77

Figure 5.2: the standard deviations of SCAM4 simulated clouds (from top to bottom: cloud fraction, LWC and IWC) due to different background data (left), error covariance matrices (middle) and precipitation (right)……...…….78

Figure 5.3: the same as figure 4 but for SCAM5……..…….79

Figure 5.4: (middle and right) model bias of cloud properties at SGP site for SCAM5 (middle) and SCAM4 (right). Blue color means model underestimation while yellow/red means overestimation, with black lines circle out where model bias is larger than the total uncertainty from large-scale forcing data and observations. (left) vertical profiles averaged in March 2000 IOP. Black line is MICROBASE, red dashed line is SCAM4, and blue dotted line is SCAM5. From top to bottom: cloud fraction, LWC and IWC……80

Figure 5.5: ensemble averaged simulations (red lines) of (a, b) LWP, (c, d) TOA net longwave radiative flux, (e, f) total cloud fraction and (g, h) surface downward shortwave radiative flux from SCAM5 (left) and SCAM4 (right) and observations from ARMBE (black lines). The red dashed lines show the standard deviation due to different large-scale forcing data……..…….…81

Figure 5.6: 4-year (July 2006 to June 2010) mean profiles of cloud fraction (right), LWC (middle) and IWC (right) at SGP site or the nearest grid. Blue line is from CAM5 GCM, red line is from C3M merged satellite retrieval, black line is from MICROBASE ground retrieval……..……82

Figure 5.7: 4-year (July 2006 to June 2010) mean total high (top), middle (middle) and low (bottom) cloud fraction in CAM5 GCM (left) and C3M merged satellite retrieval (right)…….……83

Tables

Table 3.1: different background data, error covariance matrices and constraint variables used as inputs in ensemble 3DCVA……..….……63

Table 3.2: the six background data used in ensemble 3DCVA. All datasets are interpolated into 0.5 °×0.5 °, 27 levels and 3-hourly…….……..64

Table 5.1: physical schemes used in SCAM4 and SCAM5…….……..76

Acknowledgments

My deepest gratitude goes to my dissertation advisor, Prof. Minghua Zhang, for his guidance, support and encouragement throughout these years. He has been teaching me how to strengthen myself in both ability and personality to be an independent researcher. I have benefited a lot from his physical ideas and profound knowledge. Without his time and advice this thesis would have not been accomplished.

I greatly appreciate Dr. Shaocheng Xie from Lawrence Livermore National Laboratory, and Dr. Jun Huang from the Department of Applied Mathematics and Statistics (AMS), Stony Brook University, for their help of sharing their experiences on variational analysis. Their experience and useful comments and discussions helped me a lot on the improvement of the 3DCVA algorithm. I greatly appreciate Dr. Xiangmin Jiao from AMS, Stony Brook University who gave me valuable suggestions to solve the large sparse linear system with better numerical accuracy.

I am also grateful to Dr. Ping Liu from SoMAS, Stony Brook University and Dr. Wuyin Lin from Brookhaven National Laboratory for their long time help on super-computer related issues, model-related questions and improving my scientific ideas. Sincere thanks go to my colleagues, Parama Mukherjee, Xin Xie, Tingyin Xiao, Jingyi Chen, Haiyang Yu and Xiaoxi Zhu in my group, and Dr. Yunyan Zhang from Lawrence Livermore National Laboratory for their useful discussions.

Finally, I would like to express my deepest gratitude to my wife, Jingyi Chen, my parents and my parents-in-law, for their endless love and support through my life.

Vita and Publications

Vita

2003-2007 B.S. in atmospheric sciences, Peking University, Beijing, China

2007-2010 M.S. in meteorology, Peking University, Beijing, China

2010-2015 Ph.D. in atmospheric sciences, Stony Brook University, Stony Brook, NY, USA.

Publications

Shuaiqi Tang, Minghua Zhang and Shaocheng Xie, 2015: Sensitivity Study of Ensemble Three-Dimensional Constrained Variational Analysis and Single-Column Model Test. To be submitted to *J. Geophys. Res.*

Shuaiqi Tang and Minghua Zhang, 2015: Three Dimensional Constrained Variational Analysis of Atmospheric Diabatic Heating and Derivative Fields during an ARM SGP Intensive Operational Period. submitted to *J. Geophys. Res.*

Chapter 1. Introduction

Global Climate Models (GCM) provide us a relatively complete description of the climate system and are widely used to forecast global and regional climate in future years, decades, and centuries. However, GCMs still poorly simulate certain phenomena such as Madden-Julian Oscillation (e.g. Zhang 2005), monsoons (e.g. Annamalai et al. 2007), the Inter-Tropical Convergence Zone (ITCZ) (e.g. Lin 2007) and the El-Niño Southern Oscillation (ENSO) (e.g. AchutaRao and Sperber 2006). Although computing power is growing rapidly, the resolution of major GCMs is still too coarse to simulate certain physical processes such as convections, turbulence, and cloud macro- and micro-physics. To represent these physical processes, parameterizations are always needed, which are considered as major error sources of GCMs. Therefore, it has been a subject of intensive research in the last several decades to develop sophisticated parameterizations to reduce model errors to the greatest extent possible.

One approach to improve physical parameterizations is using Large-Eddy Simulations (LES) (e.g. Khairoutdinov and Kogan 2000; Zhang et al. 2012), Cloud-Resolving Models (CRM) (e.g. Khairoutdinov and Randall 2003; Ran and Li 2013), and Single-Column Models (SCM) (e.g. Del Genio et al. 2005; Xie et al. 2005) as tools. LES and CRM can improve our understanding of certain physical processes such as cloud and turbulence because of the finer resolutions. The information from LES and CRM is then simplified and used to design GCM parameterizations. SCM is a helpful tool to simulate one column of a GCM to test the physical parameterizations against observations. This

strategy to use LES/CRM/SCM to develop and test physical parameterizations in GCM is illustrated in Figure 1.1. Observations gathered globally can be used to evaluate the performance of GCM simulations. Those intensive observations from specific field campaigns and/or observational sites can be used to derive large-scale dynamics forcing LES/CRM/SCM to develop and test physical parameterizations specifically. The large-scale dynamics prescribe the mean condition of the atmosphere and determine if the convection could be trigged, how strong or how weak the turbulence would be, how the current meteorology condition will evolve in the future, etc. In LES and CRM, some statistical quantities of clouds like the net cloud mass flux, hydrometeor mixing ratio and cloud fraction are strongly constrained by the prescribed large-scale forcing data (Khairoutdinov and Randall 2003). Thus, obtaining accurate large-scale forcing data are important for realistic LES/CRM/SCM simulations. "Derived fields" such as vertical velocity and advective tendencies are especially important but difficult to determine because they are not directly measured and involve horizontal derivatives of the mean flow in calculation which could contain large errors.

Although it is possible to directly obtain large-scale forcing data from numerical weather prediction (NWP) or reanalysis products with global coverage, parameterization deficiencies in the operational models can strongly affect the forcing data by introducing large errors (e.g. Morrison and Pinto 2004; Xie et al. 2003). Past forcing data are therefore primarily derived by careful objective analysis of atmospheric sounding data (e.g. Lin and Johnson 1996a; Ooyama 1987; Schumacher et al. 2007). Many of the observational data are obtained from the Atmospheric Radiation Measurement (ARM) program. ARM program is funded by the U.S. Department of Energy (DOE) focused on

developing highly instrumented ground stations to study clouds, aerosols and radiative transfer for more than two decades. It has high-density, multi-source surface measurements in some main sites like the Southern Great Plains (SGP), the North Slope of Alaska (NSA), and the Tropical Western Pacific (TWP) sites. These sites are typically a couple hundred kilometers across, representing an atmospheric column in a GCM. Also, there are many intensive field campaigns operated at ARM sites, which measure high-frequency surface and upper-air atmospheric quantities. These intensive measurements provide us a great platform to calculate more accurate large-scale forcing data.

Zhang and Lin (1997) developed a constrained variational analysis algorithm (hereafter 1DCVA) to improve the accuracy of atmospheric forcing data by enforcing the consistency of the atmospheric transport with auxiliary measurements at the surface and the top of the atmosphere (TOA). This method has been used in the ARM program and several other field campaigns (e.g. Schumacher et al. 2007; Schumacher et al. 2008; Xie 2005; Xie et al. 2006; Xie et al. 2010; Zhang et al. 2001) for more than one decade. It calculates the time varying vertical profiles of large-scale forcing data for an atmospheric column by adjusting state variables (horizontal wind u, v, water vapor mixing ratio q, dry static energy s. refers as background data hereafter) from sounding measurements to satisfy the conservations of mass, moisture, heat, and momentum. To ease the requirement of this method on high-density sounding measurements which are only available during specific intensive operational periods (IOPs), and to reduce errors in the NWP-derived forcing data, Xie et al. (2004) combined NWP background data with observed constraint variables to obtain long-term forcing data (the ARM Continuous

Forcing Product). Moreover, Huang (2012) made improvements to this algorithm by adding the hydrostatic balance constraints in each vertical layer and incorporating vertical and temporal correlations of errors in the 1DCVA. These algorithms so far have been designed for a one-dimensional atmospheric column representing the horizontal average of a spatial domain, which represents a GCM grid box of a few hundred kilometers across. As climate models evolve towards higher spatial resolutions, there is the need for resolution-dependent forcing data to test and improve physical parameterizations in these models. Xie et al. (2014) derived single-column forcing data for different domain sizes centered at the SGP central facility. However, for a small-size domain, surface observations are severely under-utilized. The first objective of this thesis is to extend the 1DCVA method into 3-Dimensional Constrained Variational Analysis (hereafter 3DCVA) at higher spatial resolution with additional features to improve the data quality, and to develop an ensemble framework to investigate the sensitivity of large-scale forcing data to different inputs of 3DCVA and quantify the uncertainties.

As shown in Yanai et al. (1973), the large-scale forcing data can be used to calculate the Q1 (apparent heating sources) and Q2 (apparent moisture sinks). Q1 and Q2 have been widely used to reveal the role of diabatic heating in atmospheric processes such as the Madden-Julian Oscillation (MJO) (e.g. Lin et al. 2004; Yanai et al. 2000), monsoons (e.g. Garcia and Kayano 2011; Hung and Yanai 2004), energy and moisture budget (e.g. Katsumata et al. 2011; Lin and Johnson 1996b; Yang and Smith 1999) etc. Neglecting ice processes, the equations of Q1 and Q2 can be written as:

$$\mathrm{Q1} = \frac{\partial \bar{s}}{\partial t} + \bar{\vec{V}} \cdot \nabla \bar{s} + \bar{\omega} \frac{\partial \bar{s}}{\partial p} = Q_{rad} + L_v \left(c - e \right) - \frac{\partial \overline{\omega' s'}}{\partial p} \tag{1.1}$$

$$\mathrm{Q2} = -L_v\left(\frac{\partial \bar{q}}{\partial t} + \bar{\vec{V}}\cdot\nabla\bar{q} + \bar{\omega}\frac{\partial \bar{q}}{\partial p}\right) = L_v\left(c-e\right) + L_v\frac{\partial \overline{\omega' q'}}{\partial p} \qquad (1.2)$$

Where $s = C_p T + gz$ is the dry static energy; q is vapor mixing ratio; $\vec{V}$ is the horizontal wind in vector form; L_v is the latent heat of vaporization; Q_{rad} is radiative heating rate; c is condensation rate; e is evaporation rate; ω is vertical velocity in pressure coordinate; p is pressure; the overbar refers to horizontal average in each grid (omitted in later equations) and the prime refers to a deviation from the average. Most of the previous research on Q1 and Q2 are focused on their 1D vertical profiles. Recently, Ling and Zhang (2013) and Wright and Fueglistaler (2013) compared the global distribution of climatological diabatic heating profiles in different reanalyses. They found that the gross features of Q1 are consistent in mid-latitude and in the middle and lower troposphere, while in the tropics and in the upper troposphere/lower stratosphere there are large differences among the different reanalysis products. As there are quite large uncertainties in the climatological means, it can be expected that regional Q1 and Q2 in the reanalyses on transient time scales would contain even larger uncertainties. With the 3DCVA developed, this thesis will present a spatial distribution of Q1 and Q2, and evaluate the corresponding Q1 and Q2 in several analysis/reanalysis products associated with a mid-latitude cyclone case.

Clouds are the largest error source in current GCM simulations. Cloud simulations not only have large discrepancies from satellite- and ground- based observations, but also differ a lot among different models. Zhang et al. (2005) found that the majority of ten GCMs only simulated 30% to 40% of the observed middle clouds and half of them underestimate low clouds comparing to satellite cloud analysis from ISCCP

(International Satellite Cloud Climatology Project) and CERES (Cloud and the Earth's Radiant Energy System). Xie et al. (2005) evaluated 9 SCMs and 4 CRMs using MICROBASE (ARM Baseline Microphysical Retrieval) ground measurements for a mid-latitude cyclone system, and found that the models generally captured the bulk characteristics of frontal clouds but differed significantly in the detailed structures. Klein et al. (2009) compared the simulations from 17 SCMs and 9 CRMs for an Arctic mixed phase cloud case. They found that the median simulated liquid water path (LWP) is about one-third of the observed value, and the spread among models are quite significant due to different physical schemes. Jiang et al. (2012) examined 19 CMIP5 (Coupled Model Intercomparison Project phase 5) models and found both model-observation difference and model spread are large especially at the upper troposphere levels. With the uncertainties from large-scale forcing data and observations specified, this thesis tries to diagnose the sources of model-observation bias and to isolate the errors from physical parameterizations.

The main objectives of this thesis are:

(1) introducing a 3DCVA approach;
(2) introducing an ensemble framework and analyzing the sensitivity of 3DCVA atmospheric derivative fields to input data;
(3) presenting a 3-dimensional structure of Q1 and Q2 associated with a mid-latitude cyclone and evaluating the corresponding analysis/reanalysis; and
(4) diagnosing the source of cloud bias in the Community Atmospheric Models (i.e. CAM4 and CAM5) by using the 3DCVA products.

This thesis is organized as below: Chapter 2 describes the method of the 3DCVA. Chapter 3 introduces an ensemble framework and investigates the sensitivities due to different inputs of 3DCVA. Chapter 4 presents a spatial distribution of Q1 and Q2 associated with a mid-latitude cyclone and evaluates the corresponding analysis/reanalysis. Chapter 5 shows a model-observation comparison and diagnoses the model bias. Chapter 6 summarizes the conclusions and discusses future work.

Chapter 2. The Method of Three-Dimensional Constrained Variational Analysis (3DCVA)

2.1 Model of the 3DCVA

The general idea of 3DCVA is to adjust the background data (horizontal wind *u*, *v*, water vapor mixing ratio *q*, dry static energy *s*) within their uncertainties to satisfy some constraint equations. Following Zhang and Lin (1997), the column-integrated conservations of mass, water vapor and heat are used as constraint equations:

$$\left\langle \nabla \cdot \vec{V} \right\rangle = -\frac{1}{g}\frac{dP_s}{dt} \tag{2.1}$$

$$\frac{\partial \langle q \rangle}{\partial t} + \left\langle \nabla \cdot \vec{V} q \right\rangle = E_s - P_{rec} - \frac{\partial \langle q_l \rangle}{\partial t} \tag{2.2}$$

$$\frac{\partial \langle s \rangle}{\partial t} + \left\langle \nabla \cdot \vec{V} s \right\rangle = R_{TOA} - R_{SRF} + L_v P_{rec} + SH + L_v \frac{\partial \langle q_l \rangle}{\partial t} \tag{2.3}$$

In the above constraint equations, the bracket represents vertical integration from the surface to the top of atmosphere (TOA); $\vec{V}$ is the horizontal wind in vector form, with its west-east and south-north components as *u* and *v*, respectively; *g* is gravity acceleration constant; P_s is surface pressure; *t* is time; *q* is water vapor mixing ratio; $s = C_p T + gz$ is the dry static energy, where C_p is the specific heat capacity at constant pressure, *T* is temperature, and *z* is geopotential height; E_s is surface evaporation rate; P_{rec} is surface precipitation rate; L_v is the specific latent heat of vaporization; q_l is cloud

liquid water content; R_{TOA} and R_{SRF} are net downward radiation flux at TOA and surface; *SH* is surface sensible heat flux. Ice processes and advections of cloud hydrometeors are neglected for simplicity. The terms in the right hand sides of the equations are treated as "truth" and referred to as constraint variables in this thesis. In reality, the constraint variables would not be error-free. The errors in the magnitude and spatial distribution of constraint variables will propagate to the entire atmospheric column. Generally speaking, the surface and TOA constraint variables would shift the entire vertical profiles more than change the shape to consist the column budget. Although the possible errors in constraint variables may bring the system away from the truth, it still ensures that the whole system is self-consistent. In the constraint of these variables, the background data are carefully adjusted; vertical velocity, advections and other variables needed for cloud models are calculated (Figure 2.1).

When calculating Q1 as the residual of large-scale dynamics (Eq. 1.1), spurious heating and cooling centers are often found at upper layers near the tropopause, where no strong diabatic heating sources or sinks should exist except radiation. At these layers, the vertical gradient of *s* is large, so it amplifies the errors of vertical velocity to the errors of Q1 via vertical advections. In the 1DCVA, it uses spatial smoothing on the background data at upper levels to remove the spurious heating and cooling centers. In the 3DCVA, we introduce another physical constraint to reduce the tropopause errors.

Q1 in the atmosphere includes latent heating, radiative heating and sub-grid scale heat transport. In the calm atmosphere without clouds, latent heating and sub-grid scale heat transport are zero (or close to zero), and the only heating source is radiative

heating/cooling. The radiative heating rate should be equal to Q1 calculated from the large-scale dynamics. Therefore, we impose a radiative constraint:

$$\mathrm{Q1} = \frac{\partial s}{\partial t} + \nabla \cdot \left(\vec{V} s\right) + \frac{\partial \left(\omega s\right)}{\partial p} = Q_{rad} \tag{2.4}$$

at each layer above observed cloud top or a specified height (400hPa in this study), whichever is higher. In this thesis, cloud top pressure data is obtained from NASA Langley (or from ARM External Value-Added Product (VAP)): Minnis Cloud Products Using VISST (Visible Infrared Solar-infrared Split Window Technique) Algorithm (Minnis et al. 2008) from satellite GOES 8, and radiative heating rate is calculated by Rapid Radiative Transfer Model for GCMs (RRTMG) using the temperature and moisture profiles from the background data. Because we do not have the cloud water and cloud ice information, we just calculate the radiative heating rate in cloud-free condition and use the radiative heating rate above the cloud top described before, so the impact of cloud top to the heating rate above it is ignored. The new constraint provides additional information to upper atmosphere and removes all spurious large heating and cooling centers.

The anomalies of atmospheric state variables are usually assumed as following Gaussian distribution. This is a reasonable assumption for *u*, *v*, and *s*, but not for *q* at upper layers. In the upper atmosphere, the value of *q* is quite small, so the Gaussian distribution can make *q* less than zero which is unphysical. To solve this problem, we rewrite the variable *q* to $e^{\ln q}$, and then Eq. 2.2 becomes:

$$\frac{\partial\left\langle e^{Q}\right\rangle}{\partial t}+\left\langle \nabla\cdot\vec{V}e^{Q}\right\rangle = E_s - P_{rec} - \frac{\partial\langle q_l\rangle}{\partial t}, \quad \text{where } Q=\ln q \tag{2.5}$$

The error of Q follows Gaussian distribution better than q, and we adjust Q as a new variable instead of q. It guarantees that q is positive definite no matter how large the adjustment of Q is.

2.2 Numerical Algorithms

The adjustments of background data are obtained by minimizing the cost function from Zhang and Lin (1997):

$$I=(u-u_o)^T B_u^{-1}(u-u_o)+(v-v_o)^T B_v^{-1}(v-v_o)+(q-q_o)^T B_q^{-1}(q-q_o)+(s-s_o)^T B_s^{-1}(s-s_o) \tag{2.6}$$

where *u, v, q, s* are column vectors of wind, water vapor mixing ratio and dry static energy for all grids in each time step. For a three-dimensional domain with I×J horizontal grids and K vertical levels, column vector $u=(u_{111},u_{211},...,u_{I11},...,u_{I21},...,u_{IJ1},...,u_{IJ2},...,u_{IJK})^T$; *v, q* and *s* are similar. The superscript *T* denotes the transpose of a vector; the subscript *o* denotes the initial state, and *B* represents error covariance matrix for each state variable. To better calculate the fluxes and improve the convergence, we use the adjusted C-grid (Figure 2.2) for background data. In the adjusted C-grid, *u* is at the center of east/west grid faces, *v* is at the center of north/south grid faces, *s* and *q* are at the center of all four grid faces and all the constraint variables (precipitation, radiation, surface fluxes) are in grid centers that represent grid averaging values.

We use the Lagrange multiplier method to minimize the cost function *I* while subject the atmospheric state variables to the constraint equations. The constraint equations (Eq. 2.1, 2.5, 2.3, and 2.4) can be written in the residual forms:

$$A_{mass} = \left\langle \nabla \cdot \vec{V} \right\rangle + \frac{1}{g}\frac{dP_s}{dt} \tag{2.7}$$

$$A_{water} = \frac{\partial \left\langle e^Q \right\rangle}{\partial t} + \left\langle \nabla \cdot \vec{V} e^Q \right\rangle - E_s + P_{rec} + \frac{\partial \left\langle q_l \right\rangle}{\partial t} \tag{2.8}$$

$$A_{heat} = \frac{\partial \left\langle s \right\rangle}{\partial t} + \left\langle \nabla \cdot \vec{V} s \right\rangle - R_{TOA} + R_{SRF} - LP_{rec} - SH - L\frac{\partial \left\langle q_l \right\rangle}{\partial t} \tag{2.9}$$

$$A_{radiation} = \frac{\partial s}{\partial t} + \nabla \cdot \left(\vec{V} s \right) + \frac{\partial \left(\omega s \right)}{\partial p} - Q_{rad} \tag{2.10}$$

So the 3DCVA model can be described as minimizing *I* while keeping all *A*=0. Since the units of these residuals of constraints are different, we normalize these constraint equations by dividing the equations by the horizontal mean standard deviations (std.) of the time series of the original residual *A*. Some tunable factors *C* are multiplied on the standard deviation (5 on std. of moisture constraints and 10 on std. of radiation constraints) aiming for improving convergence. These tunable factors *C* would change the adjustments of *u, v, q, s* by telling the model to satisfy some constraints more strictly but satisfy some others less strictly. The residuals of the constraint equations after adjustment need to be analyzed to choose appropriate factors *C*.

We define $X = (u-u_o, v-v_o, q-q_o, s-s_o)^T$ represents the adjustment vector of all background data, and rewrite the cost function I as:

$$I(X) = X^T B^{-1} X \tag{2.11}$$

Error covariance matrix B specifies how the variable X is adjusted. Theoretically, B should be related to the difference between the first guess (background data) and the truth, but the truth is never known. The 1DCVA uses the time variance of the background data plus instrument and measurement uncertainty to estimate the error covariance matrix, for a given background variable i and vertical level k, the diagonal value of the error covariance matrix is calculated by

$$B_{X_{ik}X_{ik}} = \mathrm{var}(X_{ik}) + \mathrm{unc}_i \tag{2.12}$$

where var(X) represents the variation of vector X, X_{ik} is a time series in each spatial point in 1DCVA, and the unc_i represents the instrument and measurement uncertainty estimates which are 0.5 m s^{-1} for winds, 0.2K for temperature and 3% for specific humidity (Zhang and Lin 1997; Zhang et al. 2001). All the non-diagonal values are 0, which assumes errors at different locations and for different variables are independent. This independent error covariance may be over-simplified. Huang (2012) calculated the correlations of the adjustments and found that there are strong correlations existing between different vertical levels, especially large positive correlations are found with the adjacent levels.

In ensemble 3DCVA with different inputs (see Chapter 3 for details), we use the anomaly covariance of different background data as the error covariance matrix. For each variable at horizontal location i,j and vertical level k, we calculate the anomaly of each background data to the ensemble mean:

$$\Delta X_{ijk}(b,t) = X_{ijk}(b,t) - \mathrm{mean}(X_{ijk}, b) \tag{2.13}$$

Where *b* denotes background data, *t* denotes time, and mean(*X*,*b*) denotes the mean value of *X* of different background data. We reshape the ΔX_{ijk} from two dimensions (*b*,*t*) into a vector, the error covariance matrix is calculated by

$$B_{X_1X_2} = \mathrm{cov}(\Delta X_1, \Delta X_2) \tag{2.14}$$

where cov(a,b) represents covariance of two vectors a and b. It removes the ensemble mean of the time variation and represents the spread of the ensemble background members. In the future it is also possible to revise the calculation of error covariance matrix not for the whole time period, but for each time steps, so the error covariance matrix would become time dependent and would have different features in different atmospheric conditions.

Since the adjustments would be largely correlated with each other especially for adjacent points, we include the horizontal and vertical correlations to the surrounding points and neighboring levels (Figure 2.3). For each variable not at the lateral and vertical boundaries, it has correlations to eight neighboring variables in horizontal and one variable at its upper level and one at lower level using Eq 2.14. We still assume that there is no correlation between different variables, and no correlation between different time steps. A sensitivity study of the impacts to the analyzed forcing data using different error covariance matrices will be addressed in Chapter 3. Thus, the error covariance matrix can be written in a bulk matrix form:

$$B = \begin{bmatrix} B_u & 0 & 0 & 0 \\ 0 & B_v & 0 & 0 \\ 0 & 0 & B_q & 0 \\ 0 & 0 & 0 & B_s \end{bmatrix} \tag{2.15}$$

We introduce a Lagrange multiplier λ, and define a new cost function J as:

$$J(X) = I(X) + \sum_k \lambda_k A_k(X) \tag{2.16}$$

subscript k represents all constraints in each grid box. The minimization problem becomes solving the following linear equations:

$$\frac{\partial I(X)}{\partial x_i} + \sum_k \lambda_k \frac{\partial A_k(X)}{\partial x_i} = 0 \tag{2.17}$$

$$A_k(X) = 0 \tag{2.18}$$

where x_i represents each variable in vector X (the same as x_j below). If we treat all λ_k as another variable similar to x_i, and let $X^* = \{X, \lambda\} = \{x_i^*\}$, we can rewrite the linear equations in a simple format:

$$\frac{\partial J}{\partial x_i^*} = 0 \tag{2.19}$$

Then, we use Newton's iteration method to solve this linear equation. The general idea solving an equation $f(x) = 0$ using Newton's iteration method is calculating the tangent line of $f(x)$ and its x-intercept to iteratively reach the point. Specifically in 3DCVA, we

make the Taylor expansion of the gradient of J at point x_{io}^*, and ignoring higher order terms, the previous equation becomes:

$$\frac{\partial J}{\partial x_i^*} = \frac{\partial J}{\partial x_{io}^*} + \sum_j \frac{\partial^2 J}{\partial x_{io}^* \partial x_{jo}^*} \delta x_j^* = 0 \tag{2.20}$$

or

$$\sum_j \frac{\partial^2 J}{\partial x_{io}^* \partial x_{jo}^*} \delta x_j^* = -\frac{\partial J}{\partial x_{io}^*} \tag{2.21}$$

In 3DCVA, Error covariance matrix B is a large, sparse matrix and could be ill-conditioned. Directly calculating the inversion of an ill-conditioned matrix contains large numerical errors, so it is needed to avoid the inversion of B. Note that the Eq. 2.21 could be rewritten as the following matrix form:

$$\begin{pmatrix} 2B^{-1} + \sum_k \lambda_k \frac{\partial A_k}{\partial x_{io} \partial x_{jo}} & \frac{\partial A_k}{\partial x_{jo}} \\ \left(\frac{\partial A_k}{\partial x_{jo}}\right)^T & 0 \end{pmatrix} \begin{pmatrix} \delta x_j \\ \delta \lambda_k \end{pmatrix} = - \begin{pmatrix} 2B^{-1}\left(x_i - x_{io}\right) + \sum_k \lambda_k \frac{\partial A_k}{\partial x_{io}} \\ A_k \end{pmatrix} \tag{2.22}$$

We multiply a matrix $\begin{pmatrix} B & 0 \\ 0 & I_0 \end{pmatrix}$ (I_0 is identity matrix) to cancel the inversion of B. After some derivation, the equation becomes

$$\begin{pmatrix} 2I_0 + B \times \sum_k \lambda_k \frac{\partial A_k}{\partial x_{io} \partial x_{jo}} & B \times \frac{\partial A_k}{\partial x_{jo}} \\ \left(\frac{\partial A_k}{\partial x_{jo}}\right)^T & 0 \end{pmatrix} \begin{pmatrix} \delta x_j \\ \delta \lambda_k \end{pmatrix} = - \begin{pmatrix} 2\left(x_i - x_{io}\right) + B \times \sum_k \lambda_k \frac{\partial A_k}{\partial x_{io}} \\ A_k \end{pmatrix} \tag{2.23}$$

We solve this linear matrix equation and update the variable x_{io}^{*}, then repeat these procedure iteratively.

The procedure of the 3DCVA is described below:

1. Input the initial guess of background data, constraint variables, cloud top pressure and radiative heating rate data. Pre-process the background data and constraint variables to the designed adjusted-C grids. The data and interpolation method used in this thesis are described in Chapter 3.
2. Set the initial values of all λs as 1.0 (the result is insensitive to the initial values of λs). Calculate the error covariance matrix.
3. Calculate 1st and 2nd derivative of *A*. ($\frac{\partial A_k}{\partial x_{io}}$ and $\frac{\partial A_k}{\partial x_{io}\partial x_{jo}}$, see Appendix 1)
4. Calculate δx_j^* by solving the Eq. 2.23 using LSQR method (described later).
5. Update X_o^* as $x_{jo}^{*(n+1)} = x_{jo}^{*(n)} + \delta x_j^*$, where *n* is iteration index.
6. Repeat step 3-5 with updated values.
7. Continue the loop until the norm of the vector $\left\{\frac{\partial J}{\partial x_{io}^*}\right\}$ is less than a certain threshold. Considering the convergence speed and the accuracy of the algorithm, we use 0.01 as the default threshold, and the iteration stops when reaching maximum iteration number (20 in default). If the maximum iteration number is reached, we release the threshold to one percent of the norm of $\left\{\frac{\partial J}{\partial x_{io}^*}\right\}$ at the first iteration. If it still larger than the threshold, we consider this time step as "fail to

converge", then skip this time step without adjustment. Typically it will reach the maximum iteration number 20, but satisfy the released condition since the norm at the first iteration is quite large (usually more than 10^6 times larger than the final norm). The absolute value of the norm of $\left\{\frac{\partial J}{\partial x_{io}^*}\right\}$ will change when tuning the tunable factor *C* which changes the magnitude of constraint residuals.

8. The adjusted fields are then used to calculate the large-scale forcing data for each grid.

The Eq. 2.23 is a typical type of linear system $Ax = b$ where *A* is a large, sparse matrix. For a 3DCVA domain used in this thesis, with 10×9 horizontal grids and 27 vertical levels, A is a ~10^4×10^4 matrix. Moreover, the matrix A is ill-conditioned and the condition number increases along with the increasing grid numbers, the increasing constraint numbers, and the inclusion of the horizontal and vertical correlations. Again, it will contain large numerical errors if directly solving this ill-conditioned system by multiplying the inversion of *A*. For better accuracy, we use a LSQR method (built in function in MATLAB) to solve the sparse linear matrix equation. LSQR is a conjugate-gradient method to iteratively solve the least squares solution *x* that minimizes the norm $(b - A * x)$ (Barrett et al. 1994; Paige and Saunders 1982). The LSQR method is more accurate and faster than directly solving the equation, but it is still quite time consuming and always reach its maximum iteration number 1000.

Chapter 3. Ensemble 3DCVA and Sensitivity Study

3.1 Ensemble 3DCVA

In the 3DCVA algorithm discussed in Chapter 2, there are three important inputs that can impact the accuracy of the results: background data, error covariance matrix and constraint variables. Background data are the first guess of the 3DCVA; all the adjustments are based on the original background data. Error covariance matrix determines how the background data are adjusted. Constraint variables, which are the right hand sides of the constraint equations, control the budget of the whole system. Previous studies (Xie et al. 2004; Zhang et al. 2001) have shown that precipitation is the dominant constraint variables during precipitation periods. To reduce the impact of uncertainties in these inputs and to analyze the sensitivities, we run 3DCVA using different combinations of six background data, five error covariance matrices and three constraint variables (Table 3.1) and calculate the ensemble mean. The data used in this thesis are described below.

The six background data are obtained from reanalysis/analysis products (Table 3.2). They are the latest version of reanalyses with higher resolution and more sophisticated models and data assimilation systems comparing to the earlier generation of reanalysis data.

- ERA-interim (Dee et al. 2011) reanalysis is produced by European Centre for Medium-Range Weather Forecasts (ECMWF) to replace its previous reanalysis

ERA-40. This reanalysis uses a spectral model Integrated Forecast System (IFS, release Cy31r2) with 30 minutes time step and T255 (79km) horizontal resolution with 60 vertical model levels. The data assimilation system it uses is a four-dimensional variational (4D-Var) scheme with 12-hourly analysis cycles. The upper-level atmospheric fields (temperature, wind, humidity, etc) are assimilated separately with the analyses of near-surface parameters (2-m temperature and 2-m humidity), and are archived every 6-hourly.

- CFSR (Saha et al. 2010) reanalysis is produced by the National Centers for Environmental Prediction (NCEP). It uses a Coupled Forecast System model with a spectral atmospheric model at T382 (38km) horizontal resolution and 64 vertical levels coupled with an ocean model. The data assimilation system it uses for atmosphere is Gridded Statistical Interpolation (GSI) which is a three-dimensional variational (3D-Var) scheme, but it includes the flow dependence to the background error variances and the First-Order Time interpolation to the Observation (FOTO) to account for the time aspect.
- MERRA (Rienecker et al. 2011) reanalysis is produced by the National Aeronautics and Space Administration (NASA) focusing on the assimilation of NASA's Earth Observing System (EOS) satellite data. MERRA reanalysis uses the version 5.2.0 of the Gordard Earth Observing System (GEOS-5) atmospheric model and data assimilation system. The GEOS-5 model is a finite-volume atmospheric GCM (AGCM) in ½ °latitude by 2/3 °longitude horizontal resolution with 72 vertical levels. The data assimilation in MERRA is also 3D-Var GSI scheme with 6-hourly update cycles, but it uses an incremental analysis update

(IAU) procedure which applies the correction to the forecast model through additional tendency term. It has rain rates from SSM/I and TRMM satellite assimilated but with low weight and weak impact on the system.

- JRA-55 (Kobayashi et al. 2015) is produced by the Japan Meteorology Agency (JMA) to replace its previous reanalysis JRA-25. It uses the operational semi-implicit semi-Lagrangian global numerical weather prediction model in JMA as of December 2009 with TL319 (55km) horizontal resolution and 60 vertical levels. The data assimilation system is a 4D-Var scheme with variational bias correction (VarBC) for satellite radiances.
- NARR (Mesinger et al. 2006) is the North American Regional Reanalysis produced by NCEP as a high-resolution, high-frequency regional reanalysis. It uses NECP Eta Model and its data assimilation system R-CDAS, which is a 3D-Var system. The time interval is 3-hour, horizontal resolution is 32km and vertically it has 45 levels. One most important feature in NARR is that it assimilates the observed precipitation by converting it into latent heat, so the model precipitation during the assimilation is close to that observed.
- The Rapid Update Cycle (RUC; Benjamin et al. 2004) is the only operational regional analysis data used here as background data produced by NCEP. It is also used in 1DCVA to produce continuous forcing data (Xie et al. 2004). The analysis used in this thesis is RUC2 (implemented from year 1998 to 2002) with 40km horizontal resolution, 40 vertical levels and 1 hour assimilation frequency, and we only use the 3-hour data (00Z, 03Z, 06Z, …) for the 3DCVA. The RUC analysis also uses a 3D-Var data assimilation system.

All the background data are linearly interpolated into 0.5 °×0.5 °horizontal resolution, 3-hourly time resolution, and 27 vertical levels from 1000 hPa to 100 hPa.

The constrained variables are mostly derived from ARM surface and TOA measurements. The spatial distribution of the surface stations of ARM measurement platforms can be found in (Zhang et al. 2001), which includes the following datasets:

- Surface Meteorological Observation Stations (SMOS) measuring surface precipitation, surface pressure, surface winds, temperature, and relative humidity.
- Energy Budget Bowen Ratio (EBBR) stations measuring surface latent and sensible heat fluxes and surface broadband net radiative flux.
- Eddy Correlation Flux Measurement System (ECOR) providing in situ half-hour averages of the surface vertical fluxes of momentum, sensible heat flux, and latent heat flux.
- Oklahoma and Kansas mesonet stations (OKM and KAM) measuring surface precipitation, pressure, winds, and temperature.
- Microwave Radiometer (MWR) stations measuring the column precipitable water and total cloud liquid water.

These surface measurements are interpolated into 0.5 °×0.5 °horizontal resolution covering the SGP domain (34.5-39 °N, 95-100 °W). If there are actual measurements within the 0.5 °×0.5 °grid box, simple arithmetic averaging is used to obtain the value for that grid box. Under circumstances that multiple instruments observe the same quantities, their measurements are merged in the arithmetic averaging process with a weighting function depending on their quality. If there is no actual measurement in the grid box, the

Barnes scheme (Barnes 1964) is used with the length scale of L_x=50km, L_y=50km, and L_t=6hr to fill the missing boxes. The satellite measurements of radiative fluxes and cloud top pressure are available at NASA Langley (Minnis et al. 2008) measured by the Geostationary Operational Environment Satellite 8 (GOES8) which is already in 0.5 °×0.5 ° grid box. The precipitation rate data is from the 4-km resolution gridded precipitation products from Arkansas-Red Basin River Forecast Center (ABRFC) based on WSD-88 rain radar and gauge measurements, and is averaged into 0.5 °×0.5 °horizontal resolution and 3-hourly time resolution.

Except radiative fluxes and precipitation rate, other surface data all surfer the same problem: interpolation from several stations to regular grids. The uncertainties from the interpolation could impact the column budget significantly in the absence of radiative fluxes and precipitation. However, when there is modulate to strong precipitation, the precipitation will dominate the other constraint variables, while the uncertainties from other constraint variables become less important. In this thesis, we mostly focused on precipitation period. In the case of light precipitation or no precipitation, more accurate gridded surface data are needed for the accuracy of the column budget.

In the ensemble 3DCVA, we have found different sources of background data. However, we don't have multiple sources for constraint variables. Similar as in Xie et al. (2014), we consider a 40% fractional root-mean-square error (RMSE) for radar-based rainfall products, and use the upper bound (1.4 times estimated rain rate), lower bound (0.6 times estimated rain rate) and the estimated rain rate to form the ensemble members.

The 40% uncertainty is only considered on the magnitude of the rain rate, but the uncertainty of precipitation location is not considered.

We have two methods in hands to calculate the error covariance matrix. One method is the original algorithm (Zhang and Lin 1997; Zhang et al. 2001) in 1DCVA, which is calculated from the timeseries of the background data plus instrument and measurement uncertainties (Eq. 2.12), with no correlation. The other method is calculated on anomaly covariance of different background data in each spatial grid, with different correlation types. In this method, we can have different correlation types of no correlation, vertical correlation only, horizontal correlation only, and both horizontal and vertical correlations. In total we can have five error covariance matrices for ensemble members. Totally we have 90 ($6 \times 3 \times 5$) 3DCVA members. When testing the sensitivity to one input (such as background data), we average the 3DCVA results over the other two input datasets (such as error covariance matrix and precipitation rate).

In this thesis, we choose the ARM March 2000 intensive operational period (IOP) at the Southern Great Plains (SGP) as a case study. This IOP contains several precipitation events with various types of clouds. We apply the 3DCVA in a rectangle domain from 34.5-39 °N, 95-100 °W at SGP (hereafter SGP domain, see Figure 4.1(a)), with $0.5\,° \times 0.5\,°$ horizontal resolution and 3-hour time resolution. In this chapter, we are focusing on the SGP domain averaged large-scale forcing data during the whole IOP. The three-dimensional analysis of a mid-latitude cyclone case during this period will be addressed in Chapter 4.

Figure 3.1 shows the domain averaged precipitation and ensemble mean temperature, water vapor mixing ratio, vertical velocity (omega), horizontal and vertical advections of dry static energy (*s*, all figures on *s* and advection of *s* in this thesis are divided by specific heat capacity at constant pressure C_p so it is in the unit of K) and moisture (*q*) from ensemble 3DCVA. There are several precipitation events during March 2-4, March 8, March 11, and March 16-19, separately. The rising motion, moisture and dry static energy advections correspond well with these events. The general patterns of these forcing data are quite consistent with forcing data derived from 1DCVA (not shown), which uses the same constraint variables within a slightly smaller domain at SGP.

Figure 3.2 shows the domain averaged ensemble mean analyzed horizontal winds (*u*, *v*), vapor mixing ratio (*q*) and dry static energy (*s*) from ensemble 3DCVA, and the domain averaged ensemble mean adjustments for each variable. The mean adjustments of horizontal winds are smaller than 0.5 m s^{-1}, but the standard deviation on horizontal grids (not shown) are in the magnitude of 1 m s^{-1}, showing that the adjustments on horizontal winds have compensate effect horizontally, which bring large change to the divergence and advections. The mean adjustments of vapor mixing ratio are mostly smaller than 0.5 g kg^{-1}, except in one low-level point at March 4th, 18Z. The possible reason of this large adjustment may come from the relatively large uncertainties in surface constraint variables in the absence of precipitation. When the constraint variables are considered as the truth, the possible large uncertainties cause the imbalance of column budget which requires more moisture added on top of the relatively low vapor mixing ratio profiles in background data. The adjustments of *s* show strong diurnal cycle with

warm adjustments of ~0.3*K* in the daytime and cold adjustments of ~0.1*K* in the nighttime. The horizontal standard deviations have the similar magnitude with the adjustments. This systematic adjustment indicate that there are inconsistent diurnal cycle between background data and surface (TOA) energy budget, which may due to the unrealistic simulation of diurnal cycle in reanalysis/analysis models. The discontinuity at ~700mb and ~300mb may be related with the issue of vertical resolution changing from 25mb at boundary layer and tropopause to 50mb at mid-troposphere. In future update of 3DCVA, a constant vertical interval of 25mb will be used.

3.2 Sensitivity of Large-Scale Forcing Data

To understand the sensitivities of large-scale forcing data to different background data, error covariance matrices and constraint variables, we calculated the standard deviations of large-scale forcing data due to different inputs of 3DCVA (Figure 3.3). When we calculate the standard deviation due to one input (such as background data), we average the large-scale forcing data over the other two input datasets (such as error covariance matrix and precipitation rate). The standard deviations of the analyzed fields due to background data are shown in the left column. Those due to error covariance matrices and precipitation are shown in the middle and right columns respectively. From top to bottom, the fields are temperature, water vapor mixing ratio, vertical velocity, horizontal and vertical advections of *s* and *q*. It is seen that the magnitude of standard deviation of temperature and water vapor mixing ratio is generally smaller than 2K and 1g kg^{-1}, respectively. During the whole IOP, the mean 1-sigma uncertainty (mean standard deviation due to different inputs relative to the temporal standard deviation of

the ensemble mean field) for all the inputs is about 1.5% for temperature and 5.6% for vapor mixing ratio. If we separate precipitation periods (precip) and non-precipitation periods (non-precip), the 1-sigma uncertainty is 1.6 % (precip) and 1.4% (non-precip) for temperature and 6.3% (precip) and 4.9% (non-precip) for vapor mixing ratio. For vertical velocity and advections (sum of horizontal and vertical advections) of s (q), the magnitude of standard deviation is generally smaller than 0.2Pa s^{-1} and 10K day^{-1} (5g kg^{-1} day^{-1}), respectively. The mean 1-sigma uncertainty for all the inputs is about 21.6% (whole period), 26.7% (precip) and 16.7% (non-precip) for vertical velocity, 27.0% (whole period), 31.9% (precip) and 22.4% (non-precip) for advections of s, and 24.9% (whole period), 32.5% (precip) and 17.6% (non-precip) for advections of q.

The relative uncertainties of vertical velocity and advective tendencies are much larger than the relative uncertainties of temperature and water vapor mixing ratio. During precipitation period, the large-scale forcing data are more uncertain than during non-precipitation period. Advections of s at upper-levels also have large sensitivities than at mid- and low-levels, which may be related to the relatively larger vertical gradient of s. The standard deviation due to different background data is relatively larger than that due to the other two inputs in most of time, showing the importance of the initial first guess to the final forcing data. For uncertainties due to precipitation, horizontal moisture advection has much larger uncertainty than other variables, which indicates tight relationship between the horizontal moisture advection and precipitation process.

Chapter 5 will investigate the sensitivities of model simulations of clouds and radiation due to different inputs (background data, error covariance matrices and constraint variables) in producing large-scale forcing data. These sensitivities define the

range of model simulation uncertainties that can be caused by large-scale forcing data. This will enable the identification of true model errors when the simulation results are compared with observations.

Chapter 4. A Case Study of Mid-Latitude Cyclone

The major advantage of 3DCVA comparing to the original variational analysis is the three-dimensional description in the analyzed domain. In this chapter, we will use the ensemble 3DCVA results obtained in Chapter 3 to investigate the spatial distribution of diabatic heating (Q1) and drying (Q2) associated with a mid-latitude cyclone case, and evaluate the corresponding Q1 and Q2 in analysis/reanalysis products.

4.1 Synoptic Condition

We choose a mid-latitude cyclone during the ARM SGP March 2000 cloud IOP for the case study. Figure 4.1(a) shows the satellite image of the cyclone at ~00Z March 3rd, 2000 (the exact time is 23:45Z March 2nd). The red rectangle represents the SGP domain in which 3DCVA is applied. At this time snapshot, the SGP domain was under the impact of an occluded front with dry air wrapping around the cyclone from the west. A more detailed description of this frontal case can be found in an ARM-GCSS (Global Energy and Water Cycle Experiment Cloud System Study) case study by Xie et al. (2005). Figure 4.1(b) shows the ABRFC precipitation rate in the SGP domain at 00Z March 3rd interpolated into 3-hour and 0.5 °×0.5 ° resolution. This time snapshot is chosen because of the clear boundary between cloudy and clear regions that can be used to verify the results.

The synoptic structure of the 00Z March 3rd cyclone in different analysis/reanalysis is plotted in Figure 4.2, which shows the temperature (red lines),

geopotential height (black lines) and specific humidity (colors) at 850 hPa. All these datasets well capture the cyclone passing to the SGP domain. The low pressure center is located to the west of the SGP domain, with a cold center to its north or northwest. The specific humidity shows a comma-shaped moisture field, with dry air on its north and west sides. At the scale of a few hundred kilometers within the SGP domain, however, differences in the moisture field among these datasets are quite large. The differences are even more prominent in the vertical velocity field. Figure 4.3 shows the corresponding temperature (red lines), geopotential height (black lines) and vertical velocity (colors) at 500 hPa. While rising motion (cool color) dominates the SGP domain and the surrounding region to the east, and downward motion (warm color) prevails at the west, the spatial pattern and the magnitude of the vertical velocity within the SGP domain is quite different among different datasets. The large differences of vertical velocity illustrate the potential uncertainties of using these products as large-scale forcing data of SCMs, CRMs and LES.

4.2 3D Structure of Q1 and Q2

If the atmospheric analysis/reanalysis satisfy the thermodynamic and moisture continuity equations, the vertically integrated Q1 and Q2 (hereafter referred to as VIQ1 and VIQ2 respectively) should be equal to the total constraint variables in the right hand sides of Eq. 2.2 and Eq. 2.3. When the latent heating from precipitation dominates the physical processes, positive VIQ1 and VIQ2 should be good approximations of the rain rates. Negative VIQ1 implies column cooling due to radiation while negative VIQ2 implies column moistening due to surface evaporation. Figures 4.4 and 4.5 show VIQ1

and VIQ2, respectively, from (a) ensemble 3DCVA, (b) ensemble background and (c-h) different background members. Although it has been shown that the global distributions of climatological VIQ1 from reanalyses are consistent with observed rain rates (Ling and Zhang 2013), Figures 4.4 and 4.5 show that on regional synoptic scale (~5 °by 5 °domain size), none of the background data captures the pattern of the precipitation. The ensemble background data are more consistent with the precipitation pattern than any of the individual members, but there are still some biases in terms of position of maximum and minimum values. The 3DCVA forces both VIQ1 and VIQ2 to be consistent with column-integrated constraints, in this case dominated by rain rates, thus both VIQ1 and VIQ2 in 3DCVA are much closer to the observed surface precipitation.

We next examine Q1 and the associated circulation (streamlines) along the southwest-to-northeast cross section indicated by the bold black line in previous figures. At this time snapshot, clouds and precipitation are located primarily in the northeast half of the cross section. Q1 in ensemble 3DCVA is shown in Figure 4.6(a). Positive Q1 is seen in the mid-troposphere above 800 hPa to the east of 98^{o}W with a peak in between 600 hPa and 700 hPa, and negative Q1 is seen below 800 hPa. These features can be explained as latent heating from cloud condensation and cooling from evaporation of rain droplets below the clouds. The relatively weak heating at 750hPa may be due to the cooling by melting. These features are consistent with the corresponding relative humidity shown in Figure 4.7(a) as contours. Figure 4.7(a) also shows the corresponding cross section of Q2 (in colors). Q1 and Q2 are consistent with each other to the extent of possible differences contributed by the radiative cooling and sub-grid scale vertical transport of heat and moisture that will be discussed further later. The peak altitude of

Q2 is generally lower than that of Q1 from 98 ºW to 96 ºW. This may be explained by the upward transport of moist static energy. Eastward of 96ºW, the peak altitude of Q2 is higher than that of Q1. This may be explained by radiative cooling from clouds with high tops. Overall, the Q1, Q2, and relative humidity distributions are consistent with the satellite cloud image and physical intuitions of these fields in the region of broad upward motion of the cyclone.

We now examine these same fields in the background data. The Q1 in ensemble background (Figure 4.6(b)) has two major differences comparing to Q1 in 3DCVA: stronger heating between 350hPa to 600hPa to the west of 98 ºW and weaker cooling below 800hPa to the east of 97 ºW. These two features contribute to the generally larger VIQ1 value along the cross section in ensemble background comparing to that in ensemble 3DCVA (Figure 4.4(b) to 4.4(a)). The Q2 in ensemble background (Figure 4.7(b)) has stronger drying (positive) between 800hPa to 900hPa to the west of 98 ºW and stronger moistening (negative) below 800hPa to the east of 97 ºW. These two features contribute to the larger VIQ2 at the west and smaller VIQ2 at the east of the cross section in ensemble background comparing to that in ensemble 3DCVA (Figure 4.5(b) to 4.5(a)), and they are inconsistent with the satellite image in Figure 4.1(a). The individual background data have different problems while RUC seems have largest bias in this case.

4.3 Sub-grid Transport of Moist Static Energy

As in Yanai et al. (1973), if we subtract Eq. 1.2 from Eq. 1.1, we get

$$Q1 - Q2 = Q_{rad} - \frac{\partial \overline{\omega' h'}}{\partial p} \qquad (4.1)$$

where $h = s + L_v q$ is moist static energy. In strong precipitation events, radiative heating and cooling are relatively small in the lower and middle troposphere, so Q1−Q2 approximately represents the second term in the right hand side, the sub-grid scale transport of moist static energy.

In Figure 4.8, we composite the Q1−Q2 profiles related to the rain rate (approximately represented by VIQ1). The composite covers the full cyclone evolution from 18Z March 1st to 24Z March 3rd. In the events of precipitation larger than 20mm day^{-1}, it is clearly seen that the ensemble 3DCVA shows large low-level negative and high-level positive Q1−Q2, indicating vertical divergence of sub-grid scale moist static energy in the lower troposphere, and vertical convergence of sub-grid scale moist static energy in the upper troposphere. The altitude of the sign change in Q1−Q2 represents the height of maximum upward transport of moist static energy, which is at around 600hPa. The sub-grid scale vertical transport is expected of convective cells embedded within the frontal system. This vertical structure however is not clear in all operational analysis/reanalysis products. The representation of rain rate becomes weaker when VIQ1 is small, so the profiles in small VIQ1 region are not examined here. The analysis/reanalysis data have some gross feature of low-level negative and high-level positive Q1−Q2, but they all show weaker low-level sub-grid scale transport of moist static energy in small and moderate rain rate region (20-80 mm day^{-1}), and show problematic features in large rain rate region (>80mm day^{-1}). Possible causes of the model errors can come from inadequate physical parameterization of convection and adverse impact of the data assimilation process.

Chapter 5. Single-Column Model Simulations of Clouds

5.1 SCM results of cloud and sensitivities

The SCMs used here are the Community Atmospheric Model version 4 (SCAM4) and version 5 (SCAM5). The two versions of SCAM have similar dynamic core but different physics packages. The parameterization schemes used in SCAM4 and SCAM5 are listed in Table 5.1. Each member of the large-scale forcing data averaged in the SGP domain from the ensemble 3DCVA is used to drive both SCMs, and the ensemble mean simulation results are used for the analysis. The two models are reinitialized at 15Z every day from March 1st to March 19th for 36 hours without relaxation to avoid large model drift, and we choose the hour 9 to 33 of the simulations (00Z to 24Z of the next day) for analysis. Surface sensible and latent heat fluxes are prescribed from the forcing data.

The time-pressure cross sections of the ensemble mean cloud fraction, liquid water content (LWC) and ice water content (IWC) from SCAM4 and SCAM5 are shown in Figure 5.1, as well as the observations obtained from ARM Baseline Microphysical Retrieval (MICROBASE) during this period. MICROBASE is a cloud retrieval product combining multiple measurements from cloud radar, lidar, ceilometer, microwave radiometer and soundings. It is located at SGP central facility and has two versions with different time resolutions: 10-second intervals and 20-minutes average. We are using the 20-minutes averaged data that are further averaged into 3-hour interval in order to compare with SCAM simulations in this study. The assumption is that the time average of point observation would be comparable with the spatial average properties within a

domain. The models generally captured the several cloud events during the IOP, with good consistency to observations especially in cloud fraction.

The differences between models and observations will be discussed in the next section. These differences can come from three possible sources: the prescribed large-scale forcing data, the biases in the model physical parameterizations, and the observational biases from measurement and sampling errors. To evaluate the models, it is necessary to isolate the true model errors from the impact of uncertainties from the large-scale forcing data and observational biases.

The standard deviation of simulated cloud fraction, LWC and IWC due to using different inputs to produce the large-scale forcing data are shown in Figure 5.2 (for SCAM4) and Figure 5.3 (for SCAM5). The 1-sigma uncertainty of cloud properties at some time steps could reach up to more than 50% of simulated values, but the time evolution of the cloud distribution patterns is consistent with different forcing data, which implies that the amount of the cloud fraction and hydrometeors, not the occurrence of clouds, is sensitive to the initialization of large-scale forcing data. The mean 1-sigma uncertainty is about 34.1% (SCAM4) and 25.9% (SCAM5) for cloud fraction, 22.5% (SCAM4) and 17.6% (SCAM5) for LWC and 25.2% (SCAM4) and 20.6% (SCAM5) for IWC. Among the three inputs in 3DCVA, background data are still the largest contributor to cloud simulation uncertainties relative to the other two input data, which is similar to the results in Chapter 3. Overall, different inputs in deriving large-scale forcing data have considerable impacts on SCM performance of cloud simulation.

5.2 SCM Model Bias and Source Diagnostics

When we have the uncertainties from the large-scale forcing data, and assuming the uncertainties of observations, we can compare the model results and observations, and diagnose the source of model bias (discrepancy between models and observations). If the model bias is larger than the uncertainties from the forcing data and observations, the physical parameterizations should be the source of the model bias, and it can be considered as a "true" model error.

We calculate the model bias and compare it with the model uncertainties from large-scale forcing data plus uncertainties of observations. The uncertainties from large-scale forcing data are calculated by the doubled standard deviation (2-sigma uncertainty) of model results using different large-scale forcing data from all 90 3DCVA members mentioned in Table 3.1. The observation uncertainties used here are estimated as 15% for LWC, 55% for IWC and 25% for cloud fraction, respectively, based on the study by Zhao et al. (2014). These uncertainties are quantified by perturbing some inputs and parameters in cloud retrieval algorithms. Some limitations and concerns about the cloud retrievals are discussed later.

Figure 5.4 shows the model bias of cloud fraction, LWC and IWC for SCAM4 (middle) and SCAM5 (right), and the vertical profiles averaged in this period (left). The black lines circle out where model bias is larger than the uncertainties from forcing data and observations. In general, SCAM4 underestimates cloud fraction; it also underestimates low-level LWC while overestimates mid-level LWC. SCAM5 overestimates high-cloud while underestimates low-cloud, it also underestimates LWC.

Both SCAM4 and SCAM5 underestimate IWC and miss some cloud events. Most of these biases are significantly larger than the total uncertainties from large-scale forcing data and observations.

Although uncertainties of the cloud retrievals are given in the comparison, they could be underestimated. Zhao et al. (2014) pointed out that these uncertainties could be even larger when considering all assumptions and precipitating clouds. When comparing different retrieval products, the spread among them is much larger than the diagnosed uncertainty for each product (Zhao et al. 2012). Different input data, different treatments of rain contamination and mixed-phase clouds are mainly responsible for the large difference of retrieved LWC (Huang et al. 2012).

Although the large uncertainties in cloud retrievals prevent us diagnosing the source of model bias with high confidence, we can use other variables as cross references. Figure 5.5 shows the observed liquid water path (LWP), outgoing longwave radiation flux at TOA, total cloud fraction (daytime mean) and surface downward shortwave radiation flux (daily mean) obtained from ARM best estimate (ARMBE) products, where LWP is measured by microwave radiometer at the central facility of SGP, and radiation fluxes and total cloud fraction are obtained from GOES8 satellite. We take daily (daytime) averaging for the latter two variables in order to smooth out the strong diurnal cycle. All these variables are generally consistent with cloud fraction and hydrometeors during this period. For LWP, SCAM5 has much lower value than the observation at most of cloud events, while SCAM4 has comparable magnitude with the observation. This is consistent to the model biases of LWC. The averaged LWP (with standard deviation, same as below) during this period in the SCAM4 and SCAM5 are 170.0 ± 7.4 and

94.8 ± 6.9 g m^{-2} respectively, comparing to the observed 191.3 g m^{-2}. Note that the cloud retrievals of LWC use LWP measured at SGP central facility as a constraint, which is also used to produce the domain mean LWP, so the LWC and LWP are not totally independent. For longwave radiation flux at TOA, both SCAM4 and SCAM5 miss some cloud events where longwave flux decreases sharply in observation but has higher value in models. During the whole IOP, SCAM4 and SCAM5 simulated TOA longwave radiation flux of 233.3 ± 1.1 and 229.4 ± 1.1 W m^{-2} respectively, larger than observed 227.6 W m^{-2}, which is consistent with the smaller cloud liquid and ice contents in the models. However, model does not always underestimate clouds and overestimate longwave flux at TOA. When SCAM5 has more high cloud fraction than observations, it generally underestimates longwave radiation flux. For the total cloud fraction and surface downward shortwave flux, the daily averaging process has smoothed out some short-period cloud features, while during March 13^{th} to 15^{th}, when both models miss most of the cloud events, they show significant underestimation of total cloud fraction and overestimation of downward shortwave radiation flux. The consistency between biases of cloud fraction and hydrometeors with biases of the vertical integrated cloud properties and radiation fluxes gives us more confidence to attribute model biases to the deficiency of the model physical parameterizations.

5.3 GCM Verifications

The above analysis of model biases is limited to one single IOP. It is useful to diagnose the model biases in GCMs in order to understand the applicability of the SCM results in this IOP to the GCMs. In this section we use the comparison of CAM5 GCM

and satellite observations to examine if the biases of SCAM5 also exist in CAM5 GCM. The satellite retrievals used here is the merged CALIPSO, CloudSat, CERES, and MODIS satellite product C3M (Kato et al. 2011). It has global coverage from July 2006 to June 2010 for 4 years. The horizontal resolution is 2.5 °×2 °, so as CAM5 GCM results. Figure 5.6 shows the climatological mean cloud fraction, LWC and IWC at SGP site from CAM5 (blue), C3M (red) and MICROBASE (black). The differences between CAM5 GCM and MICROBASE during these 4 years are consistent with the differences between SCAM5 and MICROBASE during March 2000 IOP. However, there are also quite large differences between satellite retrievals and ground-based retrievals, especially for LWC. Comparing to the differences between observations and CAM5 GCM, the differences between the two cloud retrievals are smaller for IWC and cloud fraction. For LWC, however, the difference between C3M and MICROBASE is surprisingly large, while the CAM5 GCM is more close to the C3M profile, far smaller than MICROBASE. The large offset of different cloud retrievals is also found in previous research by Zhao et al. (2012) and Huang et al. (2012). Despite the large differences in LWC, the SCAM5 biases of the underestimation of low cloud fraction and overestimation of high cloud fraction are valid in the CAM5 GCM. The underestimation of the cloud ice is also consistent.

For the global distribution of clouds, we can only compare CAM5 with C3M. Because of the large differences of the hydrometeors between satellite and MICROBASE products, we only focus on cloud fraction here. The high-, mid- and low-cloud fractions from both CAM5 and C3M are shown in Figure 5.7. CAM5 overestimates high-cloud especially in mid- to high-latitude. For mid-cloud, it simulates more cloud in polar

region but less in the tropics. For low-cloud, CAM5 significantly underestimates cloud fraction over tropical oceans. The global mean fraction of high-, mid- and low-cloud in CAM5 is 37.96%, 26.81% and 43.75%, respectively, comparing to 30.22%, 28.37% and 50.32% in C3M. This model bias is overall consistent with SCAM5 bias at SGP as overestimation of high cloud and underestimation of low cloud.

Chapter 6. Conclusions and Future Work

6.1 Conclusions

This thesis has developed a new three-dimensional constrained variational analysis (3DCVA) method based on the original algorithm (1DCVA) developed by Zhang and Lin (1997). The 3DCVA enables the derivation of gridded atmospheric vertical velocity and advective tendencies that are intended to force SCM/CRM/LES models as well as derive Q1 and Q2 within atmospheric systems. The new method made the following improvements to the 1DCVA:

- The 3DCVA extends the original method from one atmospheric column into many columns. The constraint equations are satisfied in each column simultaneously. All columns interact with each other through lateral fluxes that are constrained by spatially distributed surface and TOA measurements.
- The incorporation of radiative constraints (Eq. 2.4) eliminates the large spurious heating and cooling centers near the tropopause.
- The 3DCVA incorporates horizontal and vertical correlations in the error covariance matrix. The new error covariance matrix describes the correlation between different spatial grids more realistically.
- The 3DCVA rewrite water vapor mixing ratio q to $e^{\ln q}$, and adjust $\ln q$ as a new variable. It guarantees a positive definite q under the assumption of Gaussian distribution.

- The new algorithm has been evaluated by using a case study during the ARM March 2000 IOP at the SGP. The distributions of the Q1 and Q2 fields are found to be consistent with the spatial distribution of satellite cloud images and surface precipitation which is not the case in all analysis/reanalysis products.

An ensemble 3DCVA is performed and the sensitivities of the large-scale forcing data due to different background data, error covariance matrices and constraint variables are reported. The ensemble approach reduces possible uncertainties of the 3DCVA large-scale forcing data due to the error from one specific data source, and provides a quantitative estimation of uncertainty of large-scale forcing data. The results of sensitivity study show that all the three inputs of the 3DCVA have considerable impact to the analyzed large-scale forcing data and the related SCM results, especially to vertical velocity, advective tendencies and model simulated cloud properties. It implies that although the atmospheric state variables are accurate enough, the derivative variables could have larger uncertainties so as model simulated clouds. Among the three inputs in 3DCVA, background data contributes to the largest sensitivity, while precipitation contributes to relatively larger sensitivity to horizontal moisture advection.

Although analysis/reanalysis data also provide 3D large-scale forcing data, this thesis shows that they contain large biases in the derivative terms (Q1, Q2) at regional scale of a few hundred kilometers. For the March 3rd 2000 mid-latitude cyclone case, the analyzed Q1 and Q2 in 3DCVA showed clear contrast between precipitation and non-precipitation regions, which are consistent with the intuitive interpretation of the diabatic heating and moistening fields for mid-latitude cyclones. While different analysis/reanalysis have different errors, they generally show weaker low-level cooling

and stronger low-level moistening, and weaker low-level upward transport of moist static energy.

With the uncertainty from the large-scale forcing data calculated and the uncertainty of observations estimated, this thesis compared the model results and observations, and diagnosed the source of model bias for ARM March 2000 IOP at SGP. Comparing single-column model results of CAM4 (SCAM4) and CAM5 (SCAM5) with MICROBASE cloud retrievals, we found that SCAM4 underestimates cloud fraction, and it overestimates mid-level LWC while underestimates low-level LWC; SCAM5 overestimates high-cloud while underestimates low-cloud, and it underestimates LWC; both SCAM4 and SCAM5 underestimate IWC and miss some cloud events. Most of these model biases are larger than the uncertainty from large-scale forcing data plus uncertainty from observations, indicating that these biases could be considered as "true" model error. Although large observation uncertainties and large offset between different cloud retrievals reduce the confidence of model bias diagnosis, these results are consistent with the model biases of LWP, total cloud fraction, surface and TOA radiation fluxes. A 4-year comparison between CAM5 GCM and C3M satellite retrievals further verifies that the SCM bias seen in March 2000 IOP at SGP site is robust climatologically and globally in terms of cloud fraction.

6.2 Future Work

Because of more grid columns, more constraint variables, and more complex covariance matrix, the convergence speed of the 3DCVA algorithm becomes much slower than that of the 1DCVA and the computational cost becomes much higher.

Further improvements are needed to optimize the 3DCVA algorithm in order to improve the computational performance. This could be done by replacing Newton's iteration method, improving numerical schemes of constraint equations, or using faster-converging numerical solver for large linear matrix equation.

Applying more physical constraints may make the final analysis more realistic. One physical constraint may be added is hydrostatic balance, which has been incorporated in 1DCVA by Huang (2012). Another possible constraint to be added is to limit relative humidity no larger than saturation (RH≤100%). Currently it is treated by setting a cap of 100% at post-processing. However, more physical constraints also imply slower convergence speed and more computational time. The cost and benefit of adding a constraint should be better understood.

The 3D derived fields of Q1 and Q2 can be used to better understand the coupling of the diabatic heating fields with dynamical circulations. One example is the MJO evolution. The ARM field campaigns at the Tropical Western Pacific (TWP) site can be used to apply the 3DCVA. The derived Q1 and Q2 can reveal the spatial heating and drying structure in different phases of MJO. This information can be used to investigate the roles of heating fields in the PBL and convective systems in the development and maintenance of the MJO, and the relationship between heating tilt and the evolution of MJO. Another example is the heating profiles for different convective systems from shallow convections to deep convections. ARM SGP site would be a preferred location for this purpose because of its dense surface measurements of fluxes providing relatively reasonable surface budget in non-precipitating shallow convection cases, although the uncertainty of the 2D gridded surface data and its impact to the heating profiles needs

future investigation. Q1, Q2 and other atmospheric variables in different types of convections can be composited to understand the relationship between convections and the large-scale properties such as mass flux.

The ultimate goal of the 3DCVA is to use resolution-flexible large-scale forcing data to drive SCM/CRM/LES to better understand cloud processes and their parameterizations. Climate modeling centers around the world are now developing resolution-dependent physical parameterizations. The 3D large-scale forcing data in different resolutions and the sensitivity study would help them better investigate the impact of model resolutions and test the resolution-dependent parameterizations. We envision at least two types of new experiments that can make use of the 3DCVA data. In one type, SCM/CRM/LES can be run in each column of the 3DCVA domain and produce simulations as a "regional model" to evaluate GCMs. In another type, SCM/CRM/LES can be run with forcing data at different resolutions as a tool to design scale-aware physical parameterizations of GCMs. The 3DCVA forcing data will also enable the SCM/CRM/LES models to be configured for domain sizes that are more compatible with available observational data such as those from the scanning radars at the SGP in the ARM program.

References

AchutaRao, K., and K. Sperber, 2006: ENSO simulation in coupled ocean-atmosphere models: are the current models better? *Clim Dyn*, **27,** 1-15.

Annamalai, H., K. Hamilton, and K. R. Sperber, 2007: The South Asian Summer Monsoon and Its Relationship with ENSO in the IPCC AR4 Simulations. *Journal of Climate*, **20,** 1071-1092.

Barnes, S. L., 1964: A Technique for Maximizing Details in Numerical Weather Map Analysis. *Journal of Applied Meteorology*, **3,** 396-409.

Barrett, R., and Coauthors, 1994: *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM.

Benjamin, S. G., and Coauthors, 2004: An Hourly Assimilation–Forecast Cycle: The RUC. *Monthly Weather Review*, **132,** 495-518.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137,** 553-597.

Del Genio, A. D., A. B. Wolf, and M.-S. Yao, 2005: Evaluation of regional cloud feedbacks using single-column models. *Journal of Geophysical Research: Atmospheres*, **110,** D15S13.

Garcia, S. R., and M. T. Kayano, 2011: Moisture and heat budgets associated with the South American monsoon system and the Atlantic ITCZ. *International Journal of Climatology*, **31,** 2154-2167.

Grenier, H., and C. S. Bretherton, 2001: A Moist PBL Parameterization for Large-Scale Models and Its Application to Subtropical Cloud-Topped Marine Boundary Layers. *Monthly Weather Review*, **129,** 357-377.

Hack, J. J., 1994: Parameterization of moist convection in the National Center for Atmospheric Research community climate model (CCM2). *Journal of Geophysical Research: Atmospheres*, **99,** 5551-5568.

Holtslag, A. A. M., and B. A. Boville, 1993: Local Versus Nonlocal Boundary-Layer Diffusion in a Global Climate Model. *Journal of Climate*, **6,** 1825-1842.

Huang, D., and Coauthors, 2012: An intercomparison of radar-based liquid cloud microphysics retrievals and implications for model evaluation studies. *Atmos. Meas. Tech.*, **5,** 1409-1424.

Huang, J., 2012: Constrained Variational Analysis Integrating Vertical and Temporal Correlations. PhD, Applied Math and Statistics, Stony Brook University.

Hung, C.-W., and M. Yanai, 2004: Factors contributing to the onset of the Australian summer monsoon. *Quarterly Journal of the Royal Meteorological Society*, **130,** 739-758.

Jiang, J. H., and Coauthors, 2012: Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations. *Journal of Geophysical Research: Atmospheres*, **117,** D14105.

Kato, S., and Coauthors, 2011: Improvements of top-of-atmosphere and surface irradiance computations with CALIPSO-, CloudSat-, and MODIS-derived cloud and aerosol properties. *Journal of Geophysical Research: Atmospheres*, **116,** D19209.

Katsumata, M., P. E. Ciesielski, and R. H. Johnson, 2011: Evaluation of Budget Analyses during MISMO. *Journal of Applied Meteorology and Climatology*, **50,** 241-254.

Khairoutdinov, M., and Y. Kogan, 2000: A New Cloud Physics Parameterization in a Large-Eddy Simulation Model of Marine Stratocumulus. *Monthly Weather Review*, **128,** 229-243.

Khairoutdinov, M. F., and D. A. Randall, 2003: Cloud Resolving Modeling of the ARM Summer 1997 IOP: Model Formulation, Results, Uncertainties, and Sensitivities. *Journal of the Atmospheric Sciences*, **60,** 607-625.

Klein, S. A., and Coauthors, 2009: Intercomparison of model simulations of mixed-phase clouds observed during the ARM Mixed-Phase Arctic Cloud Experiment. I: single-layer cloud. *Quarterly Journal of the Royal Meteorological Society*, **135,** 979-1002.

Kobayashi, S., and Coauthors, 2015: The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, **93,** 5-48.

Lin, J.-L., 2007: The Double-ITCZ Problem in IPCC AR4 Coupled GCMs: Ocean–Atmosphere Feedback Analysis. *Journal of Climate*, **20,** 4497-4525.

Lin, J., B. Mapes, M. Zhang, and M. Newman, 2004: Stratiform Precipitation, Vertical Heating Profiles, and the Madden–Julian Oscillation. *Journal of the Atmospheric Sciences*, **61,** 296-309.

Lin, X., and R. H. Johnson, 1996a: Kinematic and Thermodynamic Characteristics of the Flow over the Western Pacific Warm Pool during TOGA COARE. *Journal of the Atmospheric Sciences*, **53,** 695-715.

——, 1996b: Heating, Moistening, and Rainfall over the Western Pacific Warm Pool during TOGA COARE. *Journal of the Atmospheric Sciences*, **53,** 3367-3383.

Ling, J., and C. Zhang, 2013: Diabatic Heating Profiles in Recent Global Reanalyses. *Journal of Climate*, **26,** 3307-3325.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, **87,** 343-360.

Minnis, P., and Coauthors, 2008: Near-real time cloud retrievals from operational and research meteorological satellites. *Proc. SPIE Europe Remote Sens*, Cardiff, Wales, UK, , 710703-710703-710708.

Morrison, H., and J. O. Pinto, 2004: A New Approach for Obtaining Advection Profiles: Application to the SHEBA Column. *Monthly Weather Review*, **132,** 687-702.

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. *Journal of Climate*, **21,** 3642-3659.

Ooyama, K. V., 1987: Scale-Controlled Objective Analysis. *Monthly Weather Review*, **115,** 2479-2506.

Paige, C. C., and M. A. Saunders, 1982: LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares. *ACM Trans. Math. Softw.*, **8,** 43-71.

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. *Journal of Climate*, **22,** 3449-3469.

Park, S., C. S. Bretherton, and P. J. Rasch, 2014: Integrating Cloud Processes in the Community Atmosphere Model, Version 5. *Journal of Climate*, **27,** 6821-6856.

Ran, L., and X. Li, 2013: Sensitivity of cloud-resolving precipitation simulations to uncertainty of radiation calculation: Effects of large-scale forcing. *Quarterly Journal of the Royal Meteorological Society***,** n/a-n/a.

Rasch, P. J., and J. E. Kristj ánsson, 1998: A Comparison of the CCM3 Model Climate Using Diagnosed and Predicted Condensate Parameterizations. *Journal of Climate*, **11,** 1587-1614.

Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, **24,** 3624-3648.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, **91,** 1015-1057.

Schumacher, C., M. H. Zhang, and P. E. Ciesielski, 2007: Heating Structures of the TRMM Field Campaigns. *Journal of the Atmospheric Sciences*, **64,** 2593-2610.

Schumacher, C., P. E. Ciesielski, and M. H. Zhang, 2008: Tropical Cloud Heating Profiles: Analysis from KWAJEX. *Monthly Weather Review*, **136,** 4289-4300.

Wright, J. S., and S. Fueglistaler, 2013: Large differences in reanalyses of diabatic heating in the tropical upper troposphere and lower stratosphere. *Atmos. Chem. Phys.*, **13,** 9565-9576.

Xie, S., 2005: Simulations of midlatitude frontal clouds by single-column and cloud-resolving models during the Atmospheric Radiation Measurement March 2000 cloud intensive operational period. *Journal of Geophysical Research*, **110**.

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. *Journal of Geophysical Research*, **109**.

Xie, S., R. T. Cederwall, M. Zhang, and J. J. Yio, 2003: Comparison of SCM and CSRM forcing data derived from the ECMWF model and from objective analysis at the ARM SGP site. *Journal of Geophysical Research: Atmospheres*, **108,** 4499.

Xie, S., S. A. Klein, M. Zhang, J. J. Yio, R. T. Cederwall, and R. McCoy, 2006: Developing large-scale forcing data for single-column and cloud-resolving models from the Mixed-Phase Arctic Cloud Experiment. *Journal of Geophysical Research*, **111**.

Xie, S., T. Hume, C. Jakob, S. A. Klein, R. B. McCoy, and M. Zhang, 2010: Observed Large-Scale Structures and Diabatic Heating and Drying Profiles during TWP-ICE. *Journal of Climate*, **23,** 57-79.

Xie, S., Y. Zhang, S. E. Giangrande, M. P. Jensen, R. McCoy, and M. Zhang, 2014: Interactions between Cumulus Convection and Its Environment as Revealed by the MC3E Sounding Array. *Journal of Geophysical Research: Atmospheres***,** 2014JD022011.

Xie, S., and Coauthors, 2005: Simulations of midlatitude frontal clouds by single-column and cloud-resolving models during the Atmospheric Radiation Measurement March 2000 cloud intensive operational period. *Journal of Geophysical Research: Atmospheres*, **110,** D15S03.

Yanai, M., S. Esbensen, and J.-H. Chu, 1973: Determination of Bulk Properties of Tropical Cloud Clusters from Large-Scale Heat and Moisture Budgets. *Journal of the Atmospheric Sciences*, **30,** 611-627.

Yanai, M., B. Chen, and W.-w. Tung, 2000: The Madden–Julian Oscillation Observed during the TOGA COARE IOP: Global View. *Journal of the Atmospheric Sciences*, **57,** 2374-2396.

Yang, S., and E. A. Smith, 1999: Moisture Budget Analysis of TOGA COARE Area Using SSM/I-Retrieved Latent Heating and Large-Scale Q2 Estimates. *Journal of Atmospheric and Oceanic Technology*, **16,** 633-655.

Zhang, C., 2005: Madden-Julian Oscillation. *Reviews of Geophysics*, **43**.

Zhang, G. L., and N. A. McFarlane, 1995: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model. *Atmosphere-Ocean*, **33,** 407-446.

Zhang, M., C. S. Bretherton, P. N. Blossey, S. Bony, F. Brient, and J.-C. Golaz, 2012: The CGILS experimental design to investigate low cloud feedbacks in general circulation models by using single-column and large-eddy simulation models. *Journal of Advances in Modeling Earth Systems*, **4,** M12001.

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. *Journal of the Atmospheric Sciences*, **54,** 1503-1524.

Zhang, M. H., J. L. Lin, R. T. Cederwall, J. J. Yio, and S. C. Xie, 2001: Objective Analysis of ARM IOP Data: Method and Sensitivity. *Monthly Weather Review*, **129,** 295-311.

Zhang, M. H., and Coauthors, 2005: Comparing clouds and their seasonal variations in 10 atmospheric general circulation models with satellite measurements. *Journal of Geophysical Research: Atmospheres*, **110,** D15S02.

Zhao, C., S. Xie, X. Chen, M. P. Jensen, and M. Dunn, 2014: Quantifying uncertainties of cloud microphysical property retrievals with a perturbation method. *Journal of Geophysical Research: Atmospheres*, **119,** 2013JD021112.

Zhao, C., and Coauthors, 2012: Toward understanding of differences in current cloud retrievals of ARM ground-based measurements. *Journal of Geophysical Research: Atmospheres*, **117,** D10206.

Appendix

Appendix 1: numerical calculation of 1st and 2nd derivatives

The 1st derivative $\frac{\partial A}{\partial x_{io}}$ and 2nd derivative $\frac{\partial^2 A}{\partial x_{io}\partial x_{jo}}$ are calculated based on the differential form of constraint equations Eq. 2.7 – 2.10:

$$A_{mass} = \frac{1}{g}\sum_{k=KB}^{KT}\Delta p_k\left(\frac{u_{i+1}-u_i}{\Delta x}+\frac{v_{j+1}-v_j}{\Delta y}\right)_k + \frac{1}{g}\frac{dP_s}{dt} \quad \text{(A1.1)}$$

$$A_{water} = \frac{1}{g}\sum_{k=KB}^{KT}\Delta p_k\left(\frac{e^{Q_{mt}}-e^{Q_{m(t-1)}}}{\Delta t}+\frac{u_{i+1}e^{Q_{i+1}}-u_i e^{Q_i}}{\Delta x}+\frac{v_{j+1}e^{Q_{j+1}}-v_j e^{Q_j}}{\Delta y}\right)-E_s+P_{rec}+\frac{\partial\langle q_l\rangle}{\partial t} \quad \text{(A1.2)}$$

$$A_{heat} = \frac{1}{g}\sum_{k=KB}^{KT}\Delta p_k\left(\frac{s_{mt}-s_{m(t-1)}}{\Delta t}+\frac{u_{i+1}s_{i+1}-u_i s_i}{\Delta x}+\frac{v_{j+1}s_{j+1}-v_j s_j}{\Delta y}\right)-R_{TOA}+R_{SRF}-LP_{rec}-SH-L\frac{\partial\langle q_l\rangle}{\partial t} \quad \text{(A1.3)}$$

$$A_{radiation} = \frac{s_{mt}-s_{m(t-1)}}{\Delta t}+\frac{u_{i+1}s_{i+1}-u_i s_i}{\Delta x}+\frac{v_{j+1}s_{j+1}-v_j s_j}{\Delta y}+\frac{2\omega_k s_{km}-\omega_k s_{(k+1)m}-\omega_{k+1}s_{km}}{\Delta p_k}-Q_{rad} \quad \text{(A1.4)}$$

at the grid of *i*,*j* (see Figure 2.3) at the time step *t* and the vertical level *k* (for radiation constraint). Subscript *m* represents in the center of the grid, e.g.:

$$s_m = \frac{1}{4}\left(s_i + s_{i+1} + s_j + s_{j+1}\right) \quad \text{(A1.5)}$$

The 1st derivatives of the constraints at the grid (*i*,*j*) at the timestep *t* to the variables are:

For u and v:

$$\begin{cases}
\dfrac{\partial A_{mass,water,heat}}{\partial u_{ik}} = -\dfrac{\text{var}}{\Delta x}\dfrac{\Delta p_k}{g}, \quad \text{var} = 1, e^{Q_{ik}}, s_{ik} \quad \text{for } A_{mass}, A_{water}, \text{and } A_{heat}, \text{ respectively} \\
\dfrac{\partial A_{mass,water,heat}}{\partial u_{(i+1)k}} = \dfrac{\text{var}}{\Delta x}\dfrac{\Delta p_k}{g}, \quad \text{var} = 1, e^{Q_{(i+1)k}}, s_{(i+1)k} \quad \text{for } A_{mass}, A_{water}, \text{and } A_{heat}, \text{ respectively} \\
\dfrac{\partial A_{mass,water,heat}}{\partial v_{jk}} = -\dfrac{\text{var}}{\Delta y}\dfrac{\Delta p_k}{g}, \quad \text{var} = 1, e^{Q_{jk}}, s_{jk} \quad \text{for } A_{mass}, A_{water}, \text{and } A_{heat}, \text{ respectively} \\
\dfrac{\partial A_{mass,water,heat}}{\partial v_{(j+1)k}} = \dfrac{\text{var}}{\Delta y}\dfrac{\Delta p_k}{g}, \quad \text{var} = 1, e^{Q_{(j+1)k}}, s_{(j+1)k} \quad \text{for } A_{mass}, A_{water}, \text{and } A_{heat}, \text{ respectively} \\
\dfrac{\partial A_{radiation}}{\partial u_i} = -\dfrac{s_i}{\Delta x} \\
\dfrac{\partial A_{radiation}}{\partial u_{i+1}} = \dfrac{s_{i+1}}{\Delta x} \\
\dfrac{\partial A_{radiation}}{\partial v_j} = -\dfrac{s_j}{\Delta x} \\
\dfrac{\partial A_{radiation}}{\partial v_{j+1}} = \dfrac{s_{j+1}}{\Delta x}
\end{cases}$$

For *Q*:

$$\frac{\partial A_{water}}{\partial Q_i} = \left(\frac{e^{Q_i}}{4\Delta t} - \frac{u_i e^{Q_i}}{\Delta x} \right) \frac{\Delta p_k}{g}$$

$$\frac{\partial A_{water}}{\partial Q_{i+1}} = \left(\frac{e^{Q_{i+1}}}{4\Delta t} + \frac{u_{i+1} e^{Q_{i+1}}}{\Delta x} \right) \frac{\Delta p_k}{g}$$

$$\frac{\partial A_{water}}{\partial Q_j} = \left(\frac{e^{Q_j}}{4\Delta t} - \frac{v_j e^{Q_j}}{\Delta y} \right) \frac{\Delta p_k}{g}$$

$$\frac{\partial A_{water}}{\partial Q_{j+1}} = \left(\frac{e^{Q_{j+1}}}{4\Delta t} + \frac{v_{j+1} e^{Q_{j+1}}}{\Delta y} \right) \frac{\Delta p_k}{g}$$

For s:

$$
\begin{cases}
\dfrac{\partial A_{heat}}{\partial s_i} = \left(\dfrac{1}{4\Delta t} - \dfrac{u_i}{\Delta x} \right) \dfrac{\Delta p_k}{g} \\
\dfrac{\partial A_{heat}}{\partial s_{i+1}} = \left(\dfrac{1}{4\Delta t} + \dfrac{u_{i+1}}{\Delta x} \right) \dfrac{\Delta p_k}{g} \\
\dfrac{\partial A_{heat}}{\partial s_j} = \left(\dfrac{1}{4\Delta t} - \dfrac{v_j}{\Delta y} \right) \dfrac{\Delta p_k}{g} \\
\dfrac{\partial A_{heat}}{\partial s_{j+1}} = \left(\dfrac{1}{4\Delta t} + \dfrac{v_{j+1}}{\Delta y} \right) \dfrac{\Delta p_k}{g} \\
\dfrac{\partial A_{radiation}}{\partial s_{ik}} = \dfrac{1}{4\Delta t} - \dfrac{u_i}{\Delta x} + \dfrac{2\omega_k - \omega_{k+1}}{4\Delta p_k} \\
\dfrac{\partial A_{radiation}}{\partial s_{(i+1)k}} = \dfrac{1}{4\Delta t} + \dfrac{u_{i+1}}{\Delta x} + \dfrac{2\omega_k - \omega_{k+1}}{4\Delta p_k} \\
\dfrac{\partial A_{radiation}}{\partial s_{jk}} = \dfrac{1}{4\Delta t} - \dfrac{v_j}{\Delta y} + \dfrac{2\omega_k - \omega_{k+1}}{4\Delta p_k} \\
\dfrac{\partial A_{radiation}}{\partial s_{(j+1)k}} = \dfrac{1}{4\Delta t} + \dfrac{v_{j+1}}{\Delta y} + \dfrac{2\omega_k - \omega_{k+1}}{4\Delta p_k} \\
\dfrac{\partial A_{radiation}}{\partial s_{(k+1)}} = -\dfrac{\omega_k}{4\Delta p_k}, \quad \text{for } i, i+1, j \text{ and } j+1
\end{cases}
$$

Other 1st derivatives are all zero.

The 2nd derivatives only apply to some combination of the variables. The non-zero 2nd derivatives of the constraint at the grid (*i*,*j*) are:

$$
\begin{cases}
\dfrac{\partial^2 A_{water}}{\partial Q_i \partial Q_i} = \left(\dfrac{e^{Q_i}}{4\Delta t} - \dfrac{u_i e^{Q_i}}{\Delta x} \right) \dfrac{\Delta p_k}{g} \\
\dfrac{\partial^2 A_{water}}{\partial Q_{i+1} \partial Q_{i+1}} = \left(\dfrac{e^{Q_{i+1}}}{4\Delta t} + \dfrac{u_{i+1} e^{Q_{i+1}}}{\Delta x} \right) \dfrac{\Delta p_k}{g} \\
\dfrac{\partial^2 A_{water}}{\partial Q_j \partial Q_j} = \left(\dfrac{e^{Q_j}}{4\Delta t} - \dfrac{v_j e^{Q_j}}{\Delta y} \right) \dfrac{\Delta p_k}{g} \\
\dfrac{\partial^2 A_{water}}{\partial Q_{j+1} \partial Q_{j+1}} = \left(\dfrac{e^{Q_{j+1}}}{4\Delta t} + \dfrac{v_{j+1} e^{Q_{j+1}}}{\Delta y} \right) \dfrac{\Delta p_k}{g} \\
\dfrac{\partial^2 A_{water}}{\partial Q_i \partial u_i} = -\dfrac{e^{Q_i}}{\Delta x} \dfrac{\Delta p_k}{g} \\
\dfrac{\partial^2 A_{water}}{\partial Q_{i+1} \partial u_{i+1}} = \dfrac{e^{Q_{i+1}}}{\Delta x} \dfrac{\Delta p_k}{g} \\
\dfrac{\partial^2 A_{water}}{\partial Q_j \partial v_j} = -\dfrac{e^{Q_j}}{\Delta y} \dfrac{\Delta p_k}{g} \\
\dfrac{\partial^2 A_{water}}{\partial Q_{j+1} \partial v_{j+1}} = \dfrac{e^{Q_{j+1}}}{\Delta y} \dfrac{\Delta p_k}{g}
\end{cases}
$$

$$
\begin{cases}
\dfrac{\partial^2 A_{heat}}{\partial s_i \partial u_i} = -\dfrac{1}{\Delta x} \dfrac{\Delta p_k}{g} \\
\dfrac{\partial^2 A_{heat}}{\partial s_{i+1} \partial u_{i+1}} = \dfrac{1}{\Delta x} \dfrac{\Delta p_k}{g} \\
\dfrac{\partial^2 A_{heat}}{\partial s_j \partial v_j} = -\dfrac{1}{\Delta y} \dfrac{\Delta p_k}{g} \\
\dfrac{\partial^2 A_{heat}}{\partial s_{j+1} \partial v_{j+1}} = \dfrac{1}{\Delta y} \dfrac{\Delta p_k}{g}
\end{cases}
$$

$$\begin{cases} \dfrac{\partial^2 A_{radiation}}{\partial s_i \partial u_i} = -\dfrac{1}{\Delta x} \\ \dfrac{\partial^2 A_{radiation}}{\partial s_{i+1} \partial u_{i+1}} = \dfrac{1}{\Delta x} \\ \dfrac{\partial^2 A_{radiation}}{\partial s_j \partial v_j} = -\dfrac{1}{\Delta y} \\ \dfrac{\partial^2 A_{radiation}}{\partial s_{j+1} \partial v_{j+1}} = \dfrac{1}{\Delta y} \end{cases}$$

These individual elements are calculated and form the matrix of the 1st derivative $\dfrac{\partial A}{\partial x_{io}}$ and the 2nd derivative $\dfrac{\partial^2 A}{\partial x_{io} \partial x_{jo}}$.

List of Figures and Tables

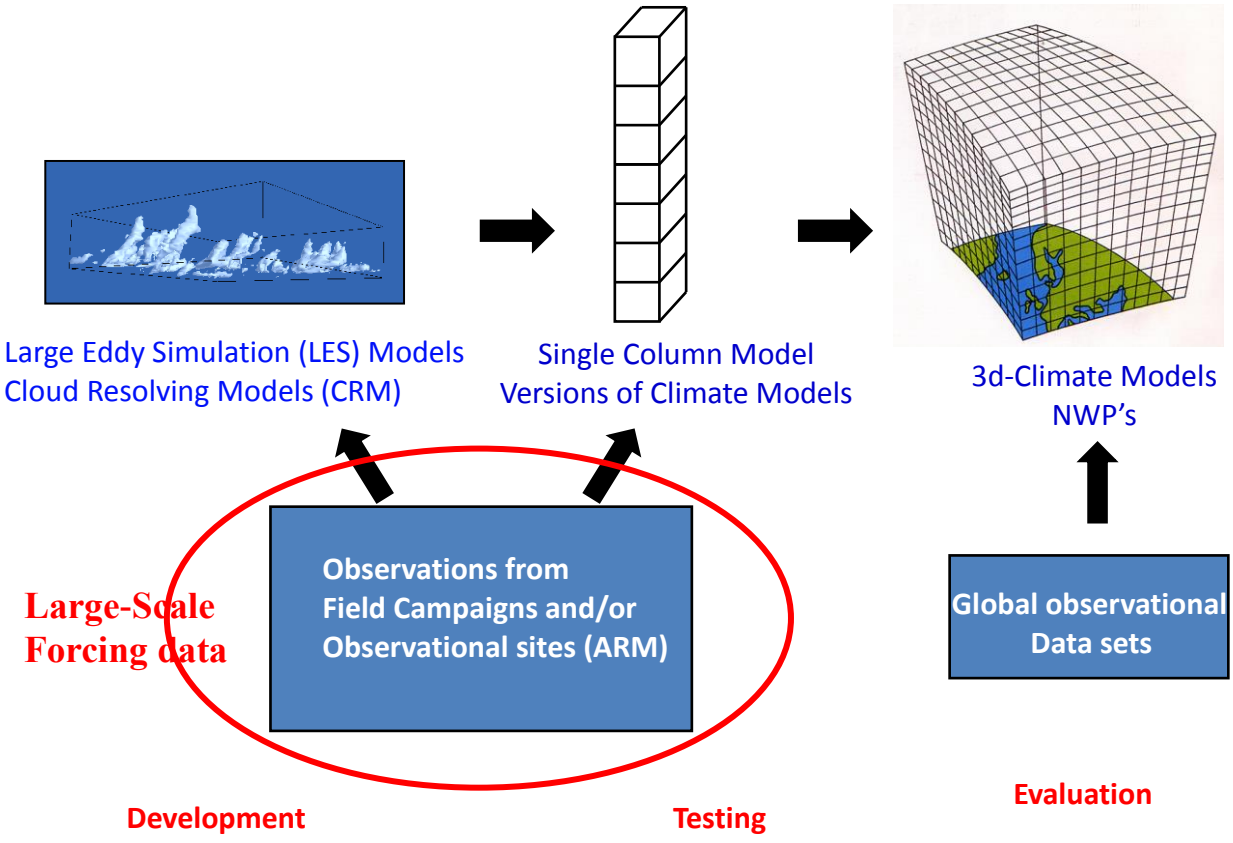

Figure 1.1: Working strategy illustration of GEWEX (Global Energy and Water Cycle Experiment) Cloud System Study (GCSS) presented by A. Pier Siebesma in 2008. Large-scale forcing data derived from field campaigns and/or observational sites are needed to drive LES, CRM and SCM which are used to develop and test the physical parameterizations in GCM and NWP.

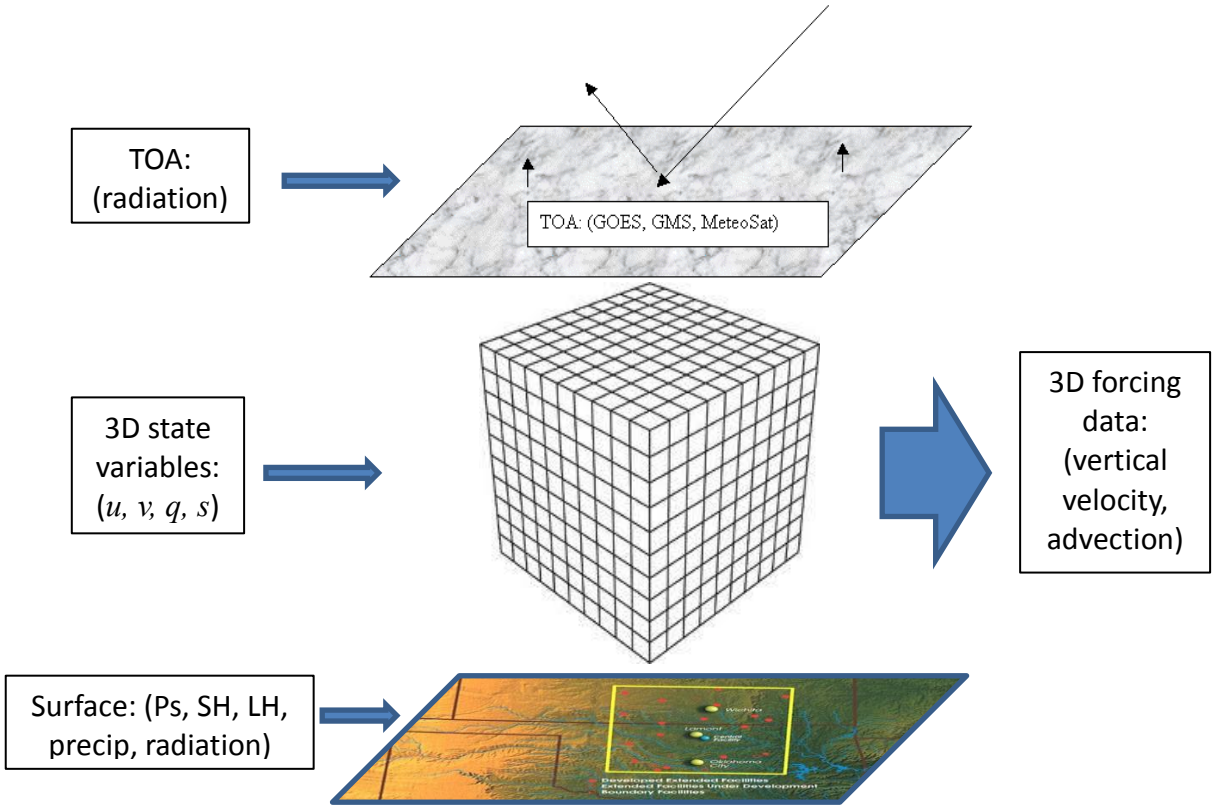

Figure 2.1: schematic figure of the 3DCVA. It shows the inputs of background data (3D state variables), surface and TOA constraint variables, and the output of 3D large-scale forcing data.

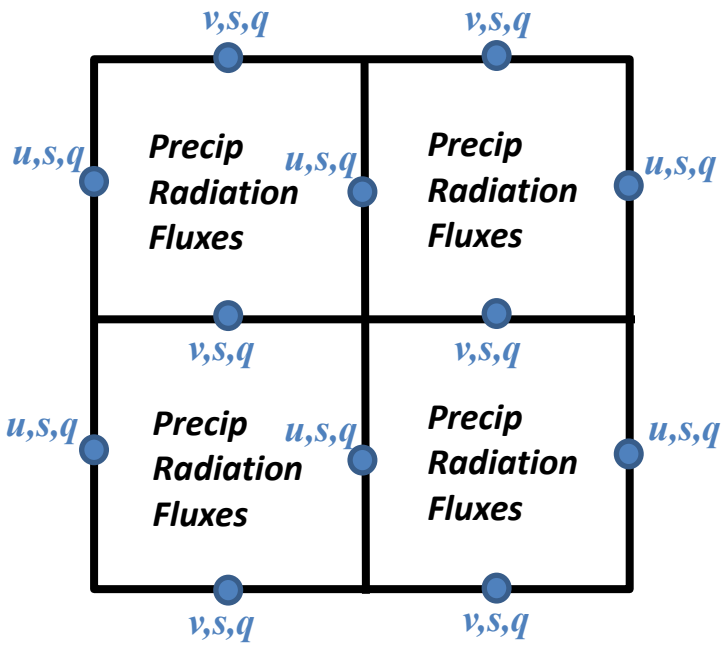

Figure 2.2: the adjusted C-grid used in 3DCVA. *u* is at the center of east/west grid faces, *v* is at the center of north/south grid faces, *s* and *q* are at the center of all four grid faces and all the constraint variables (precipitation, radiation, surface fluxes) are at the grid center and represent grid average.

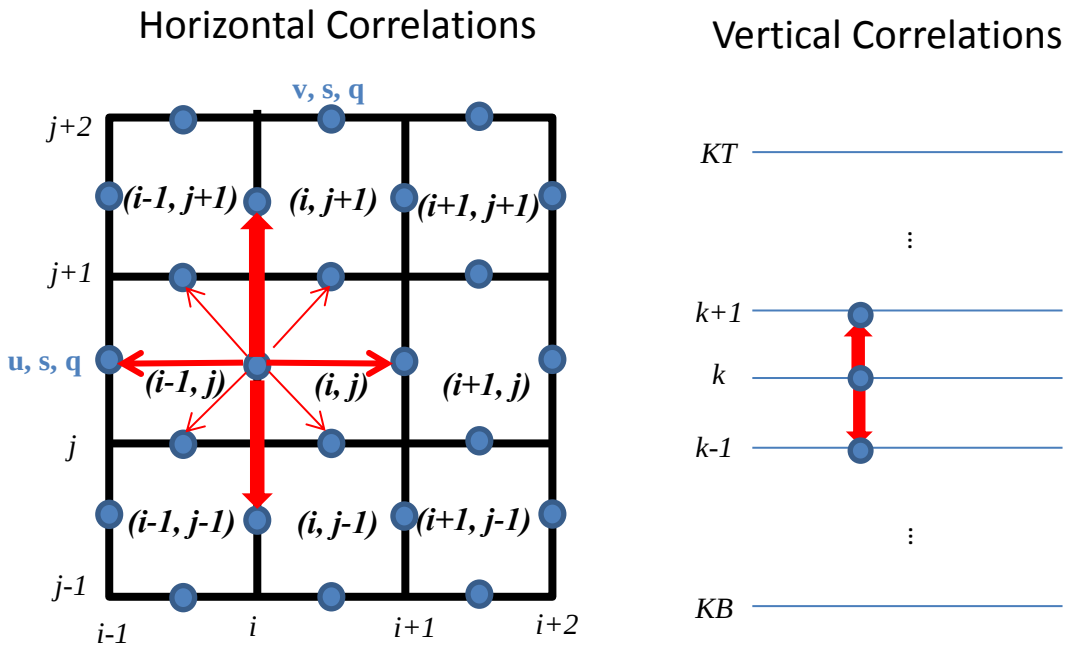

Figure 2.3: Illustration of horizontal and vertical correlations. *i, j* and *k* represent the horizontal and vertical grid/level indices. Bold ***i, j*** show the index of the grids while the thin *i, j* show the index of the boundary variables, so the grid (***i, j***) has 4 boundary variables at *i*, *i+1*, *j*, and *j+1*. *KB* is the level at the surface and *KT* is the level at the TOA. Blue dots are the variables of the background data, and the red arrows show the correlation of the center variable to its surrounding variables.

Background data	RUC, ERA-interim, CFSR, JRA55, MERRA, NARR (details see Table 3.2)
Error covariance matrix	Calculated from variance of time series, only auto correlation.
	Calculated from variance of ensemble members, only auto correlation.
	Calculated from variance of ensemble members, vertical correlation.
	Calculated from variance of ensemble members, horizontal correlation.
	Calculated from variance of ensemble members, horizontal and vertical correlation.
Constraint variables	Arkansas-Red Basin River Forecast Center gridded precipitation products
	Upper bound: 1.4 × ABRFC precip amount
	Lower bound: 0.6 × ABRFC precip amount

Table 3.1: different background data, error covariance matrices and constraint variables used as inputs in ensemble 3DCVA.

Name	Reference	Horizontal resolution	Vertical levels (1000-100hPa)	Time resolution
ERA-interim	*Dee et al.*, 2011	~0.7 °×~0.7 °	27	6 hourly
CFSR	*Saha et al.*, 2010	0.5 °×0.5 °	27	6 hourly
MERRA	*Rienecker et al.*, 2011	1.25 °×1.25 °	25	3 hourly
JRA55	*Kobayashi et al.* 2015	1.25 °×1.25 °	27	6 hourly
NARR	*Mesinger et al.*, 2006	32.463km ×32.463km	29	3 hourly
RUC	*Benjamin et al.*, 2004	40km ×40km	37	3 hourly

Table 3.2: the six background data used in ensemble 3DCVA. All datasets are interpolated into 0.5 °×0.5 °, 27 levels and 3-hourly.

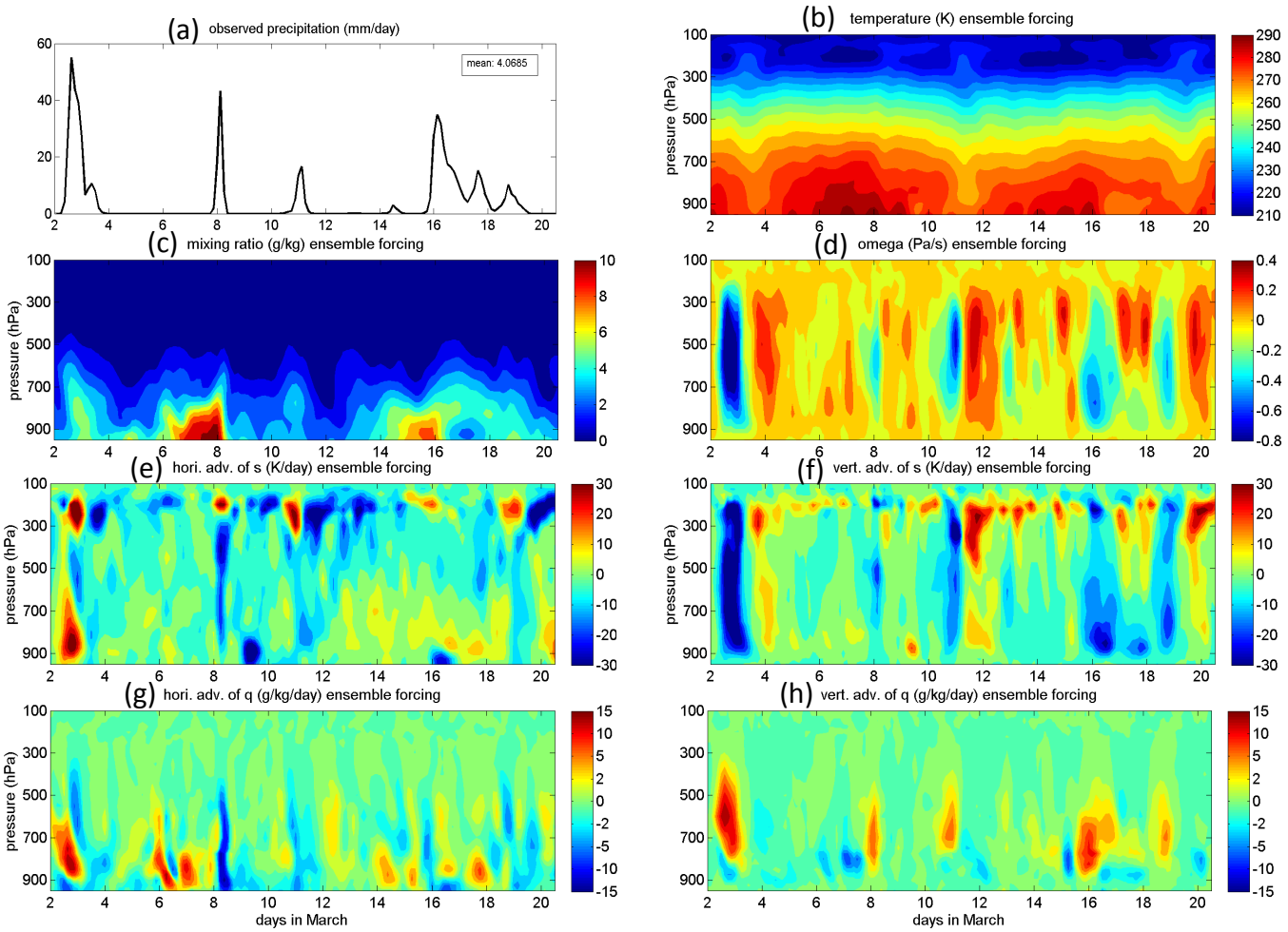

Figure 3.1: SGP domain (34.75-38.75 ºN, 95.25-99.75 ºW) averaged profile of (a) precipitation, (b) temperature, (c) water vapor mixing ratio, (d) vertical velocity, (e – h) horizontal and vertical advections of dry static energy *s* and moisture *q*.

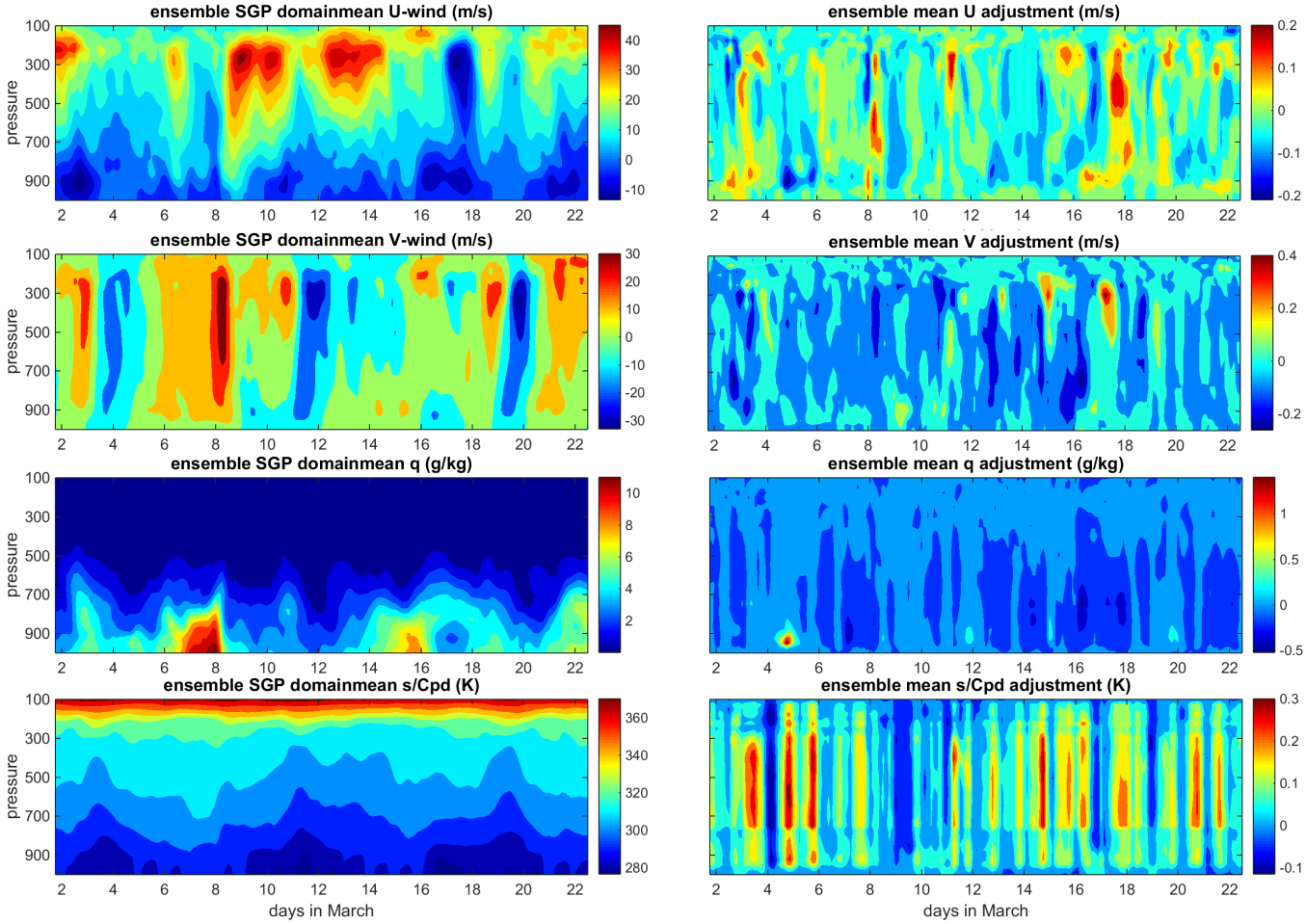

Figure 3.2: (left) ensemble mean SGP domain averaged horizontal wind, moisture and dry static energy (*u*, *v*, *q*, *s*) and (right) ensemble mean SGP domain mean adjustments of *u*, *v*, *q*, *s*.

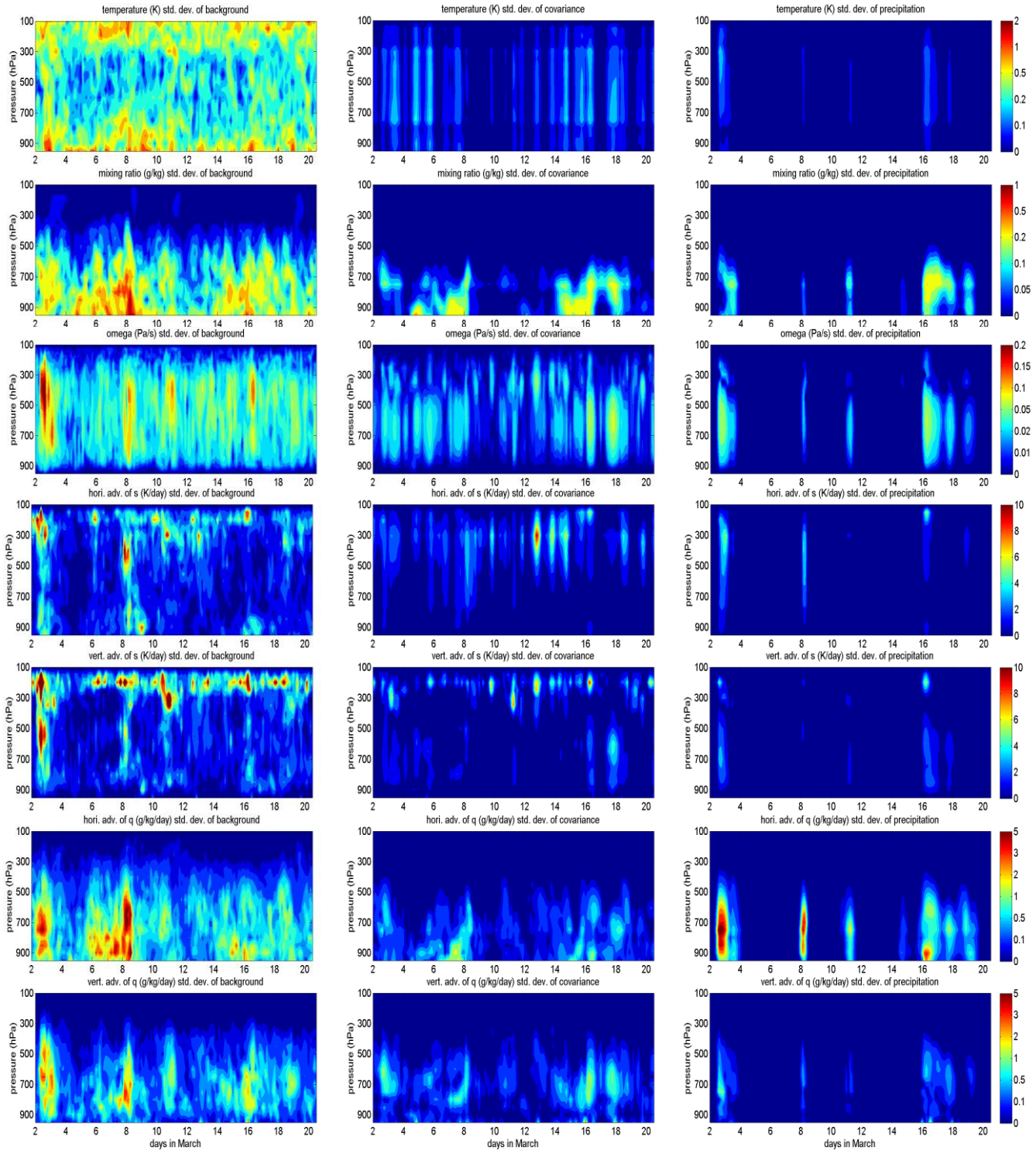

Figure 3.3: the standard deviations of large-scale forcing data (from top to bottom: T, q, ω, horizontal and vertical advections of s and q) due to different background data (left), error covariance matrices (middle) and constraint variables (right, only precipitation).

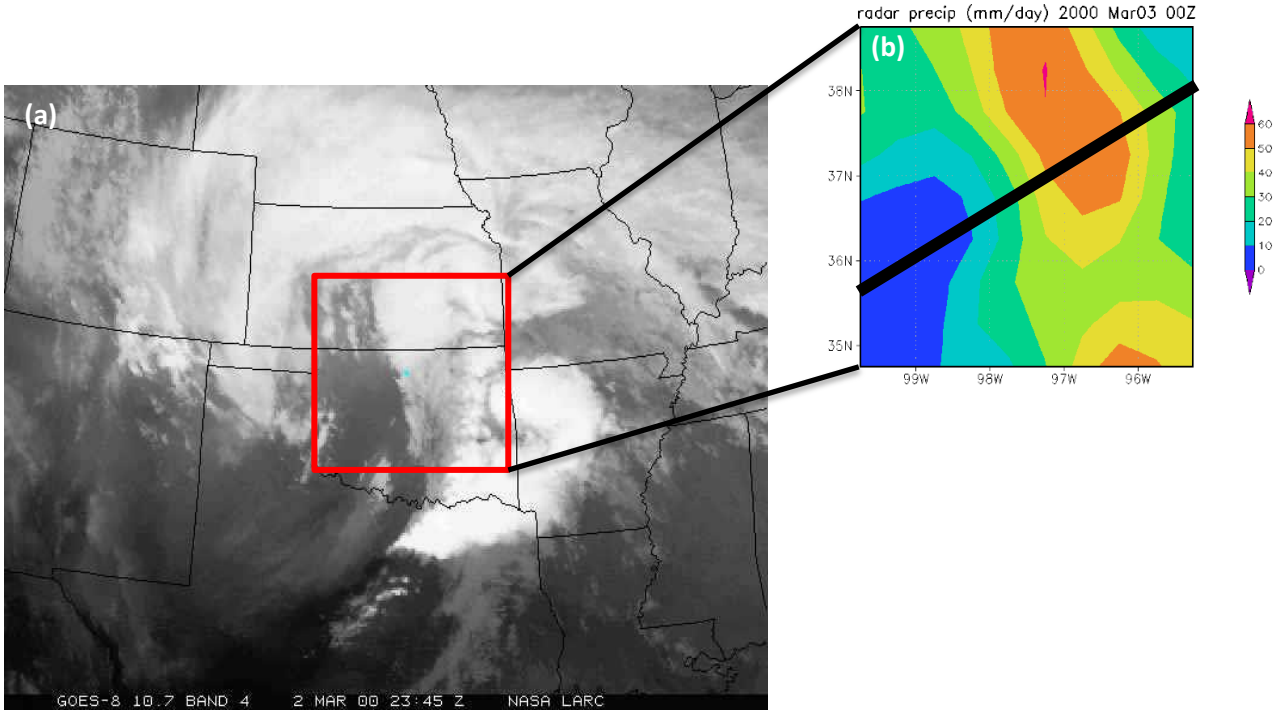

Figure 4.1: (a). GOES-8 satellite cloud image at 23:45Z March 2nd 2000. The red box shows the SGP domain applying 3DCVA. (b) 3-hourly precipitation rate centered at 00Z Mar 3rd averaged into 0.5 °×0.5 °horizontal resolution from the Arkansas-Red Basin River Forecast Center (ABRFC). The bold black line shows the position of cross-front section in later figures.

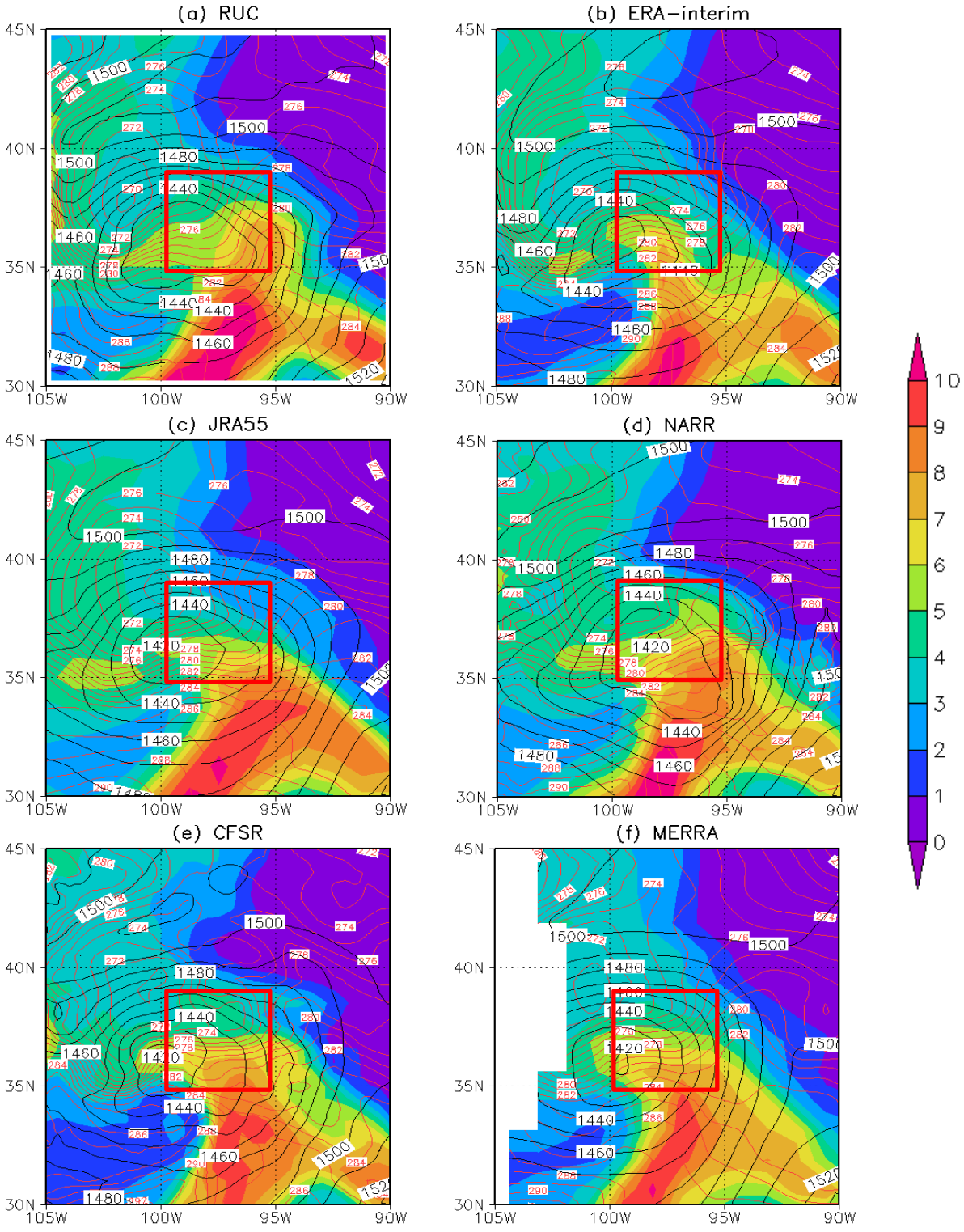

Figure 4.2: 850hPa Geopotential height (gpm, black lines), temperature (K, red lines) and specific humidity (g kg^{-1}, shaded) from (a) RUC, (b) ERA-Interim, (c) JRA 55, (d) NARR, (e) CFSR, (f) MERRA at 00Z Mar 3rd 2000. The red box indicates the SGP domain.

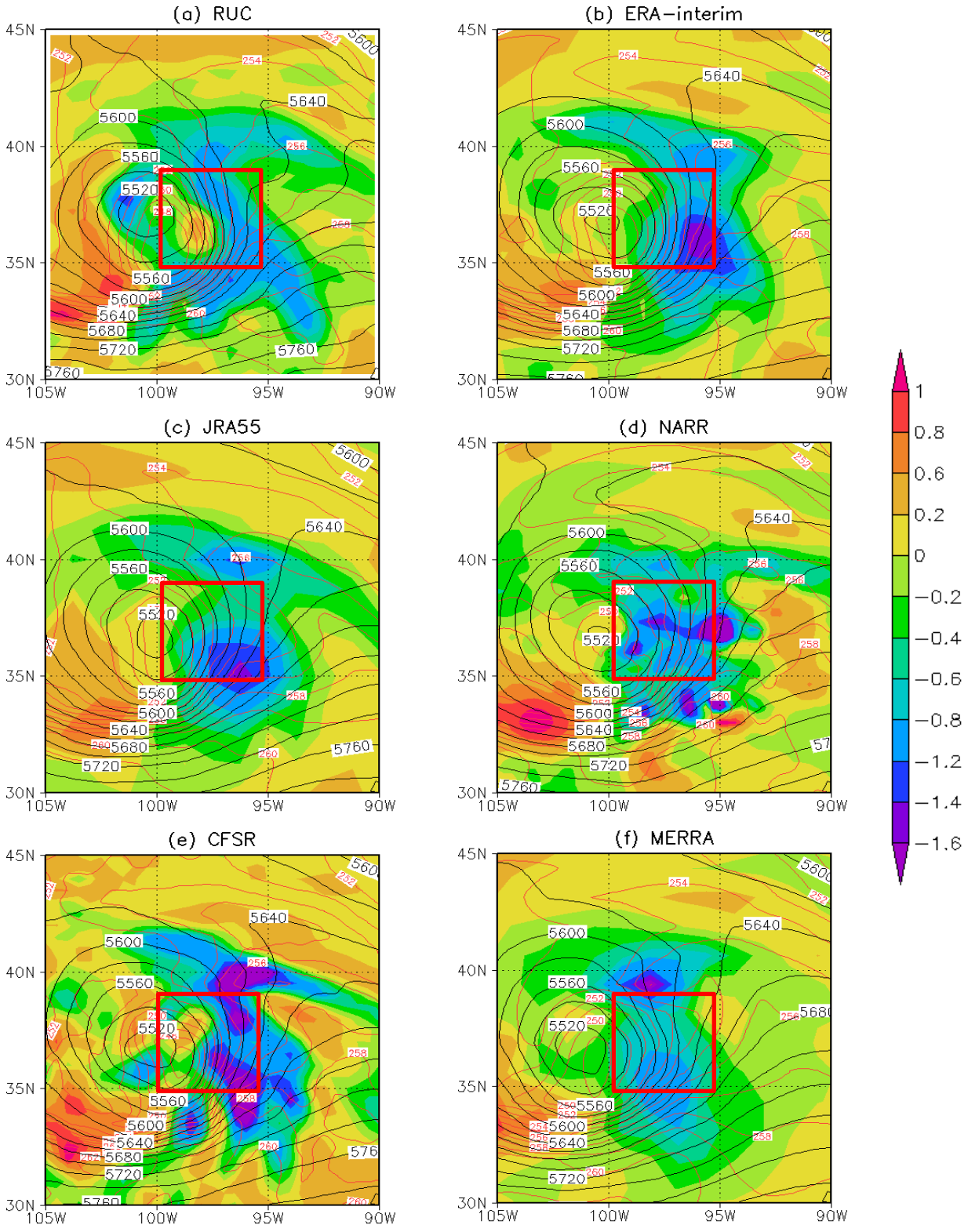

Figure 4.3: Same as Figure 4.2 except that the shading is for vertical velocity (Pa s^{-1}) and the level is 500hPa.

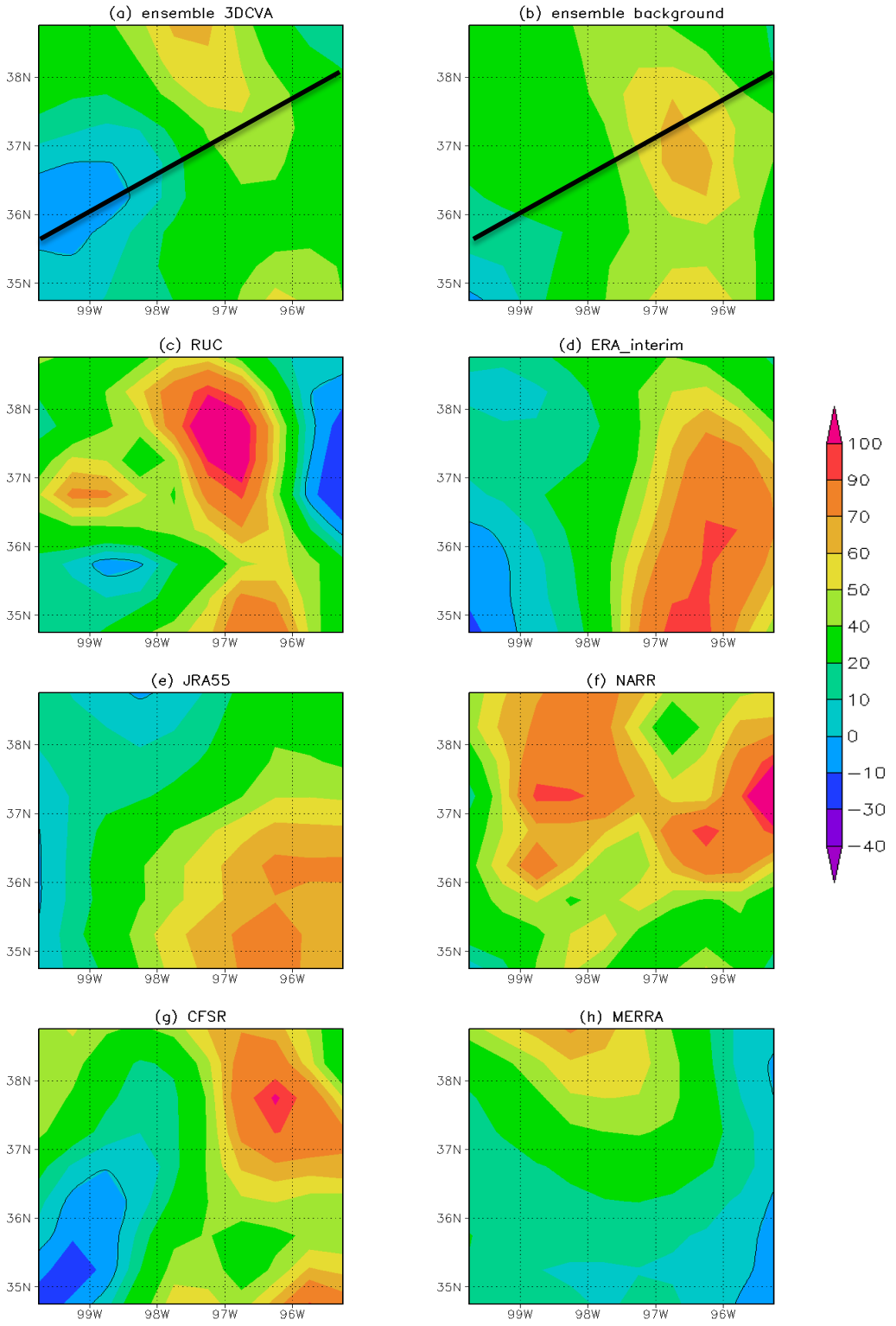

Figure 4.4: The vertically integrated Q1 from (a) ensemble 3DCVA, (b) ensemble background data, (c) RUC, (d) ERA Interim, (e) JRA 55, (f) NARR, (g) CFSR and (h) MERRA at 00Z Mar 3rd 2000. The unit is mm day^{-1}. The bold black lines in (a) and (b) shows the position of cross-front section in later figures.

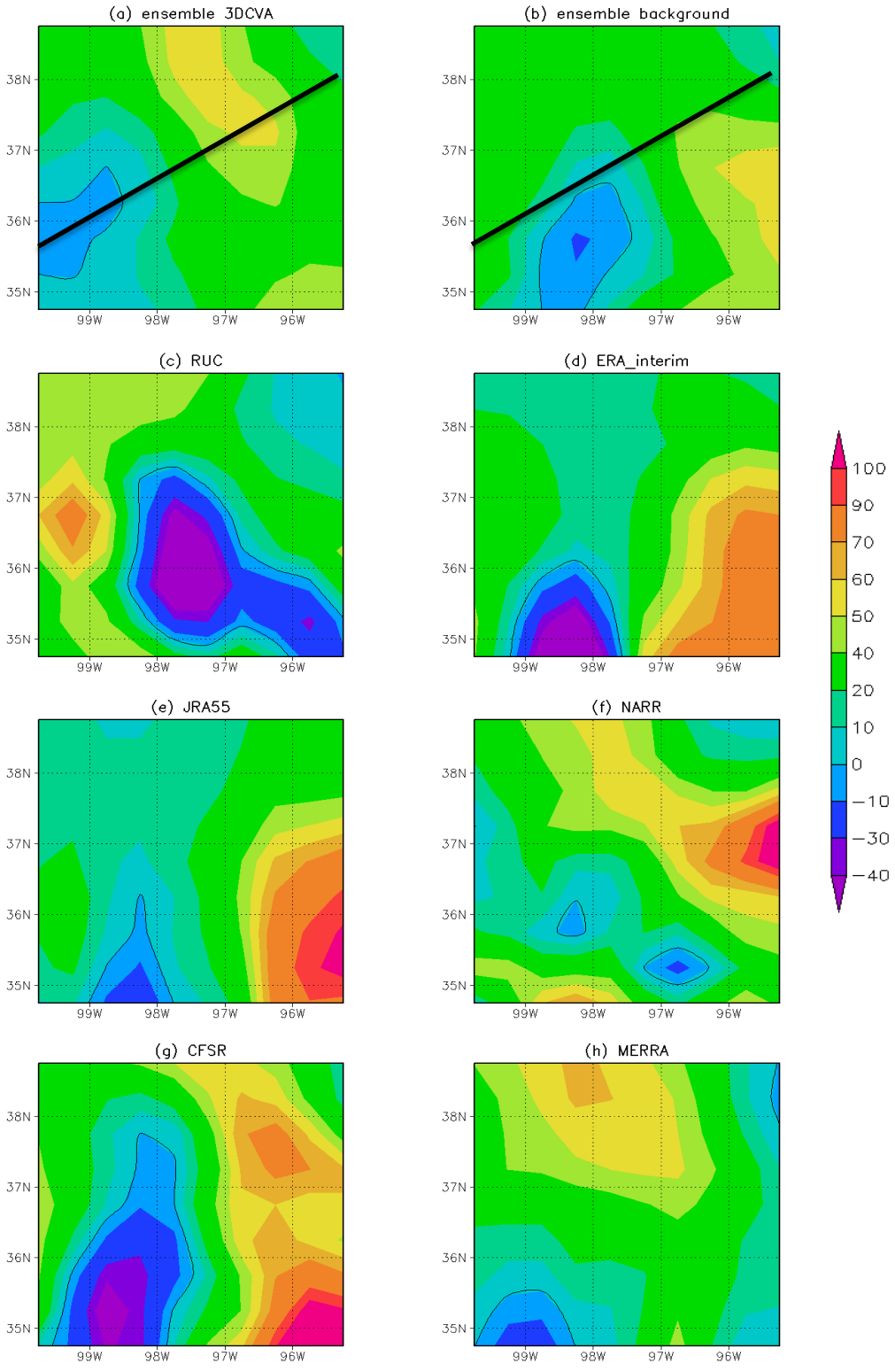

Figure 4.5: The same as figure 4.4 but for vertical integrated Q2 (mm day^{-1}).

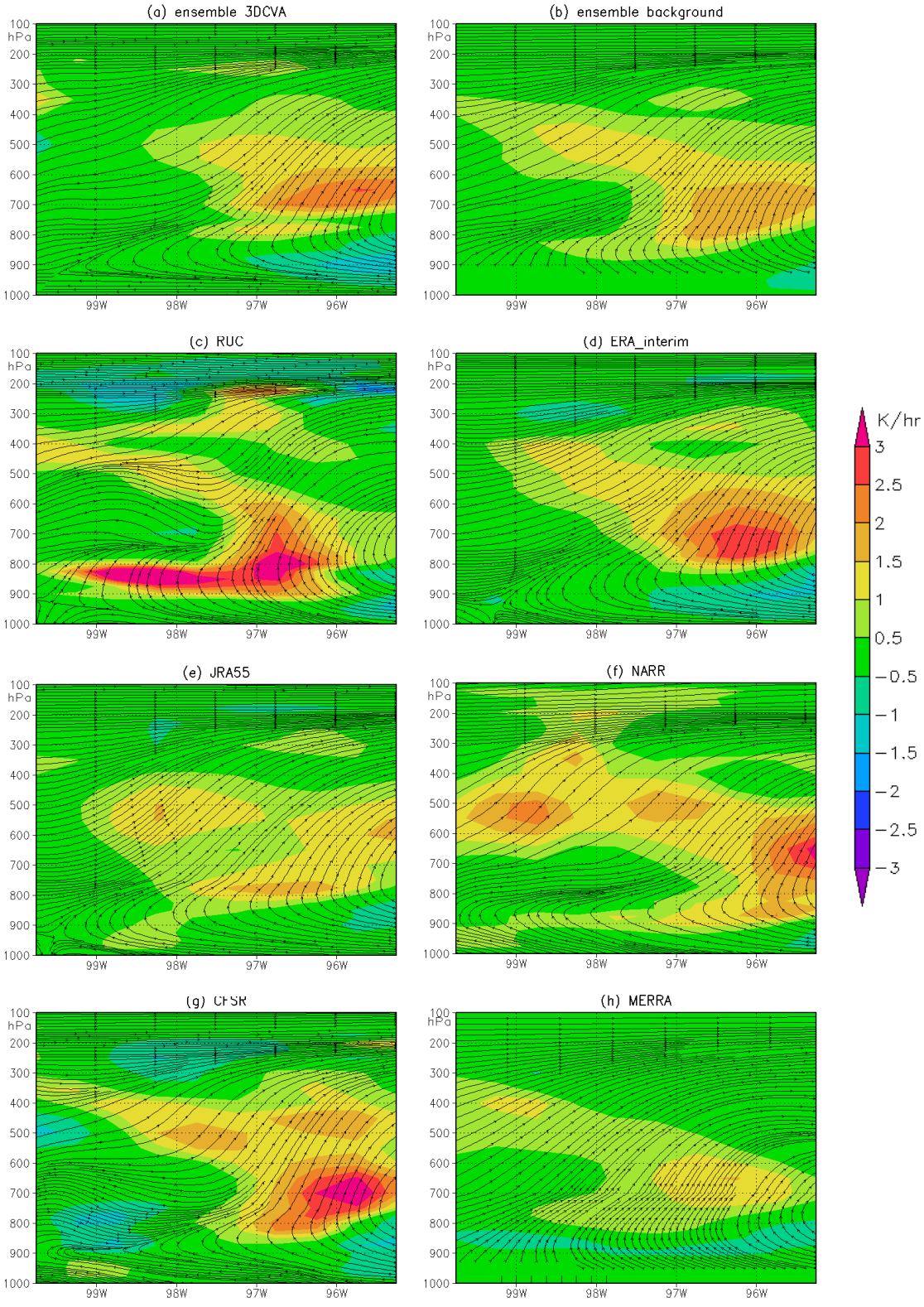

Figure 4.6: cross section (indicated by the bold black line in figure 4.1(b)) of Q1 (color, K hr^{-1}) and circulation along the cross section (streamlines) from (a) ensemble 3DCVA, (b) ensemble background data, (c) RUC, (d) ERA Interim, (e) JRA 55, (f) NARR, (g) CFSR and (h) MERRA. The time is 00Z Mar 3rd 2000. X axis shows the longitude along the cross section.

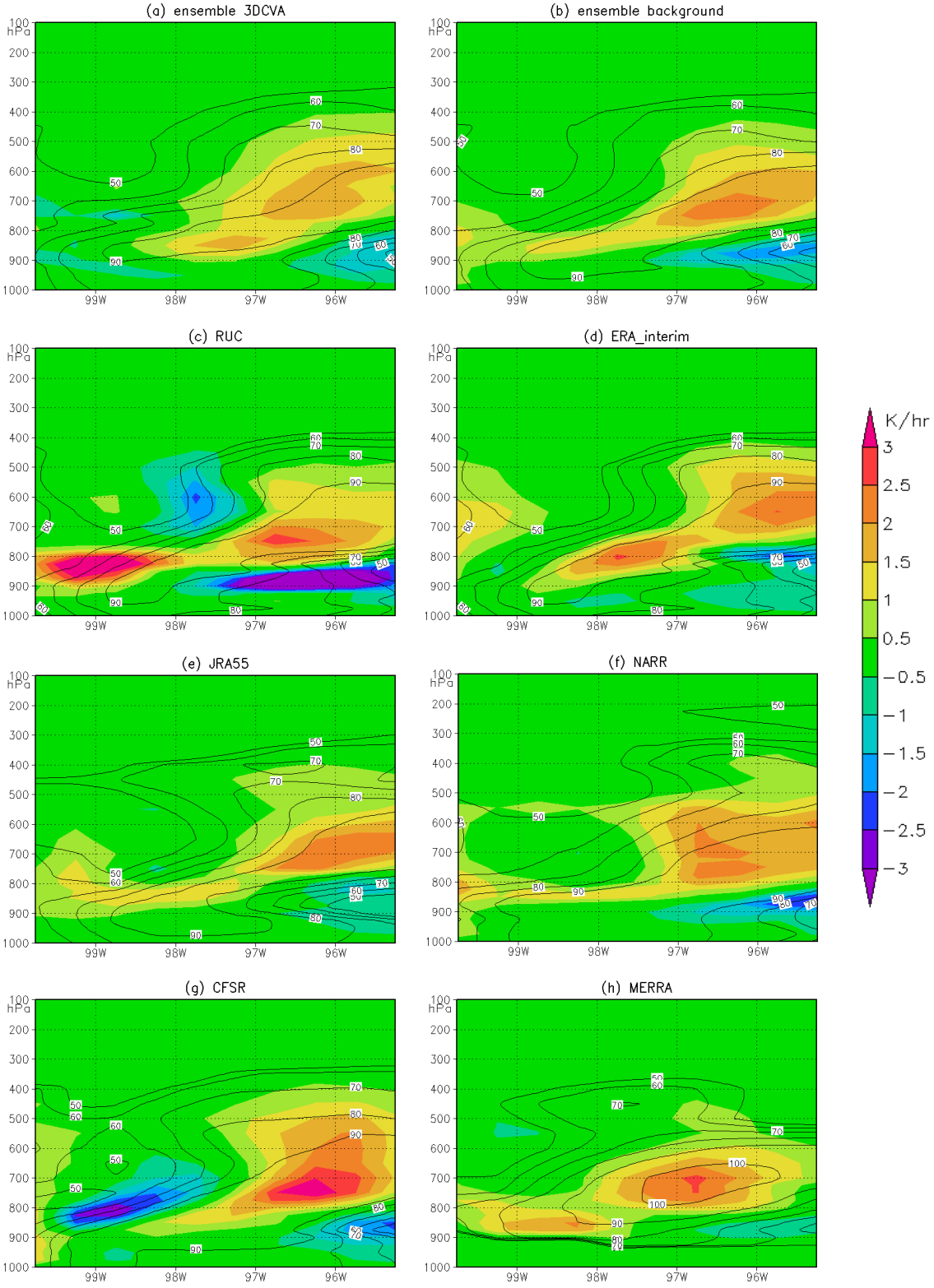

Figure 4.7: cross section (indicated by the bold black line in figure 4.1(b)) of Q2 (color, K hr^{-1}) and relative humidity (contour, %) (a) ensemble 3DCVA, (b) ensemble background data, (c) RUC, (d) ERA Interim, (e) JRA 55, (f) NARR, (g) CFSR and (h) MERRA. The time is 00Z Mar 3rd 2000. X axis shows the longitude along the cross section.

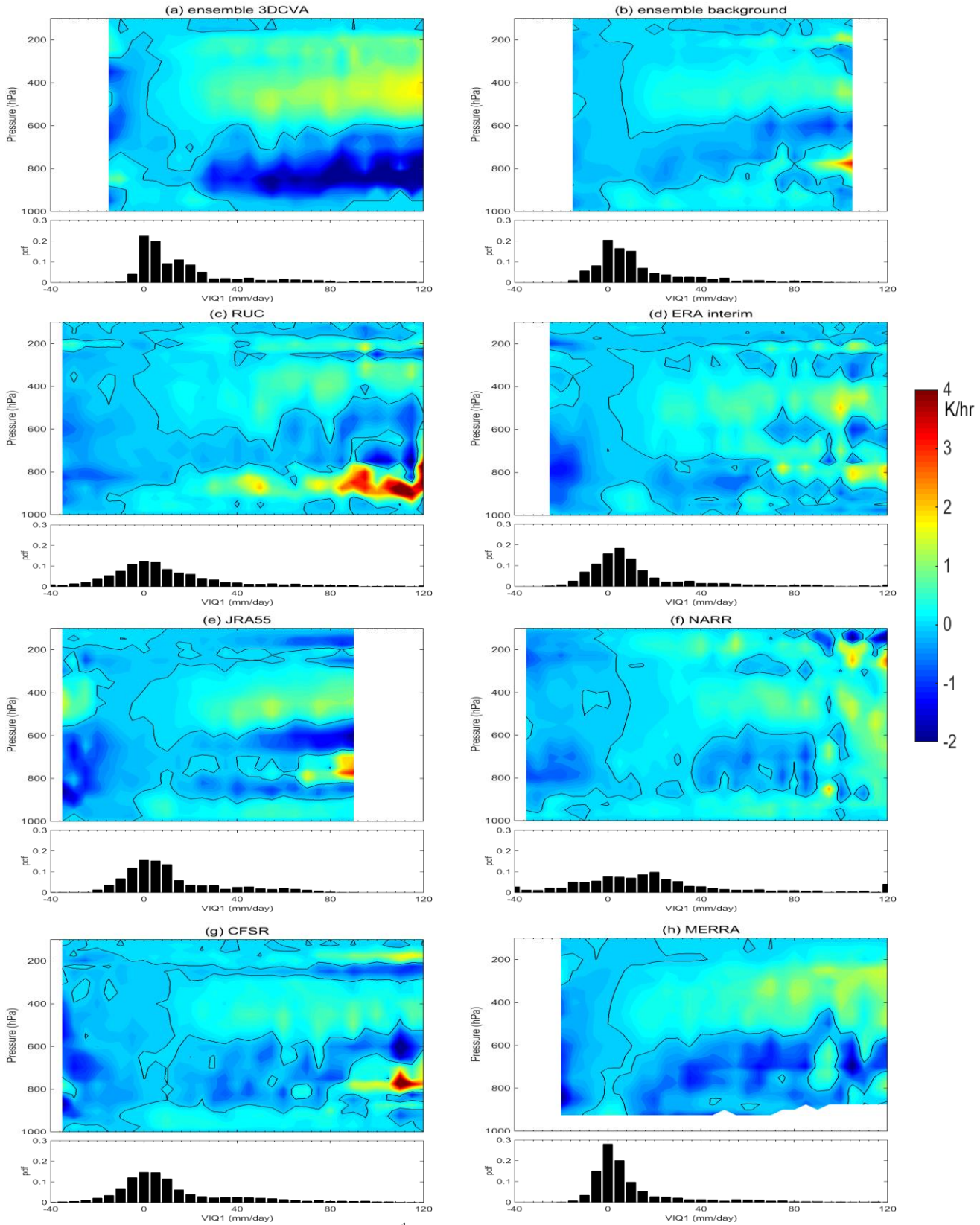

Figure 4.8: (top) Q1−Q2 profiles (K hr^{-1}) as a function of VIQ1 and (bottom) PDFs of VIQ1 (in fraction) from (a) ensemble 3DCVA, (b) ensemble background data, (c) RUC, (d) ERA Interim, (e) JRA 55, (f) NARR, (g) CFSR and (h) MERRA for all grids in SGP domain during 18Z March 1st to 24Z March 3rd.

Scheme	SCAM4	SCAM5
Eddy scheme	HB (Holtslag and Boville 1993)	Diag_TKE(Grenier and Bretherton 2001)
Micro physics	RK(Rasch and Kristj ánsson 1998)	MG(Morrison and Gettelman 2008)
Macro physics	RK(Rasch and Kristj ánsson 1998)	Park(Park et al. 2014)
Shallow convection	Hack(Hack 1994)	UW(Park and Bretherton 2009)
Deep convection	ZM(Zhang and McFarlane 1995)	ZM(Zhang and McFarlane 1995)
Radiation	CAMRT	RRTMG

Table 5.1: physical schemes used in SCAM4 and SCAM5.

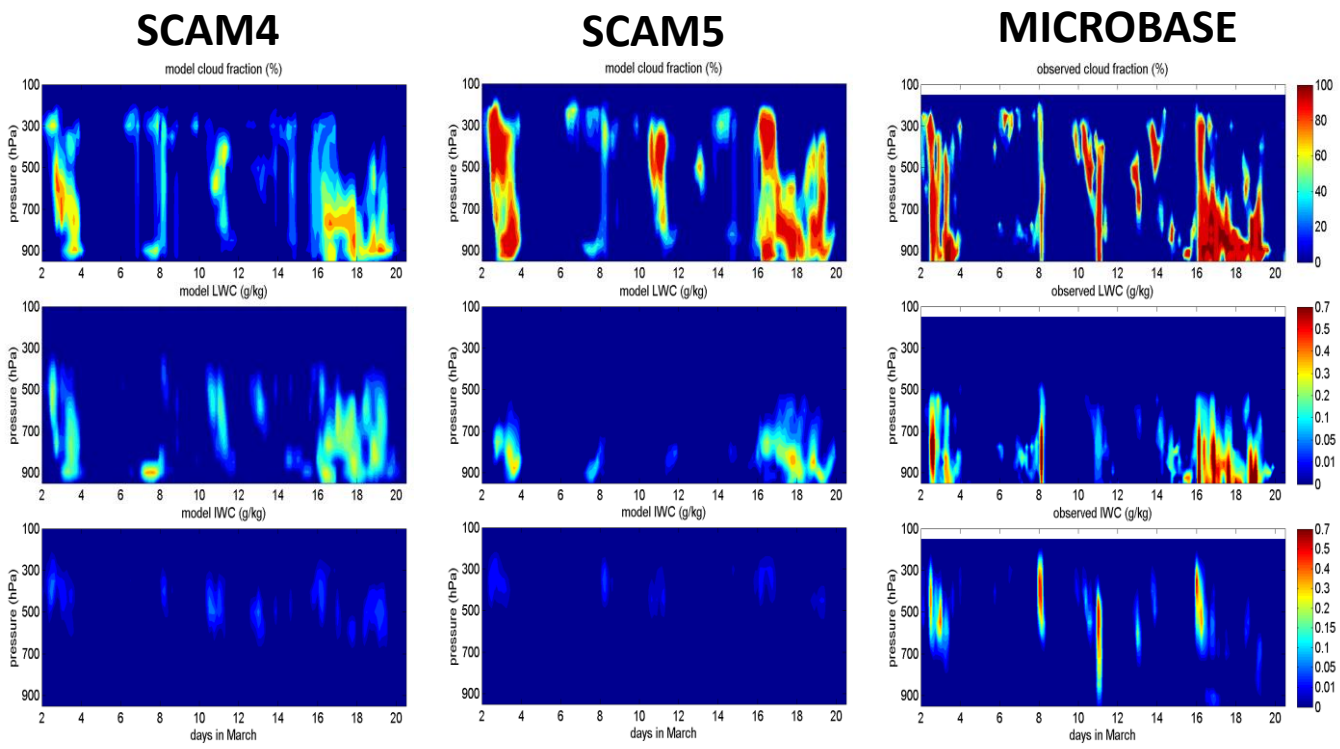

Figure 5.1: the ensemble average of simulated cloud properties in SCAM4 (left), SCAM5 (middle) and cloud retrievals from MICROBASE (right). From top to bottom: cloud fraction, LWC and IWC.

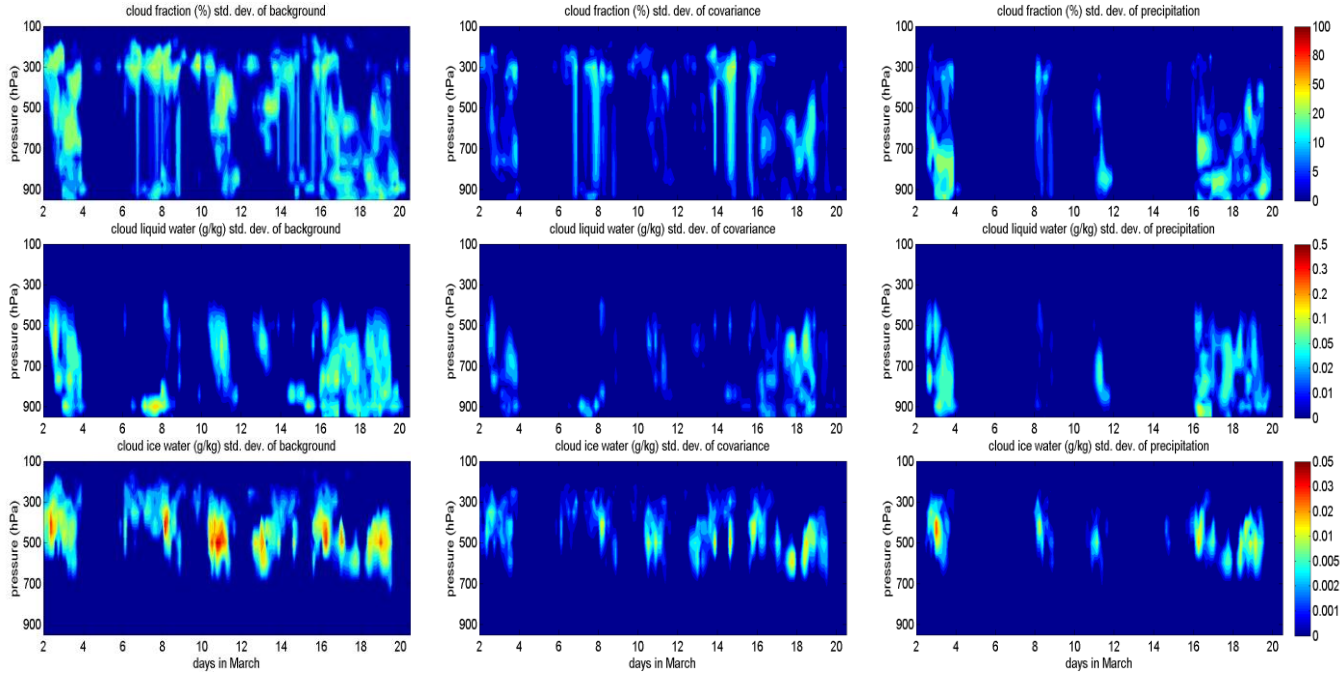

Figure 5.2: the standard deviations of SCAM4 simulated clouds (from top to bottom: cloud fraction, LWC and IWC) due to different background data (left), error covariance matrices (middle) and precipitation (right).

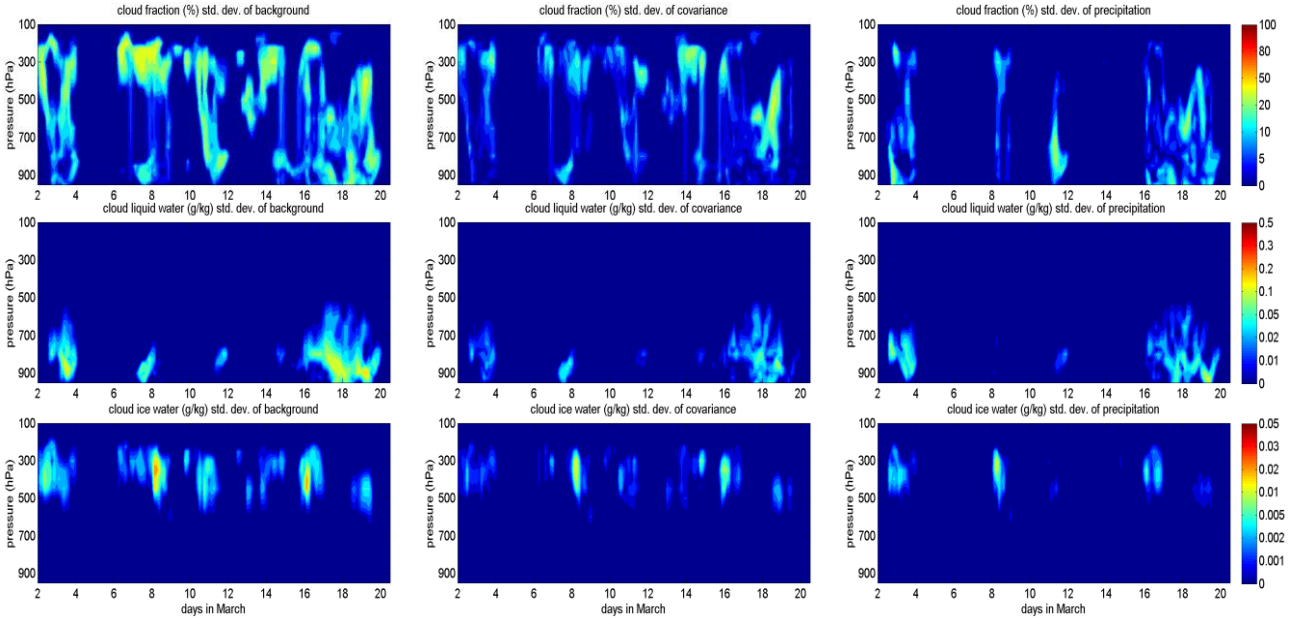

Figure 5.3: the same as figure 4 but for SCAM5

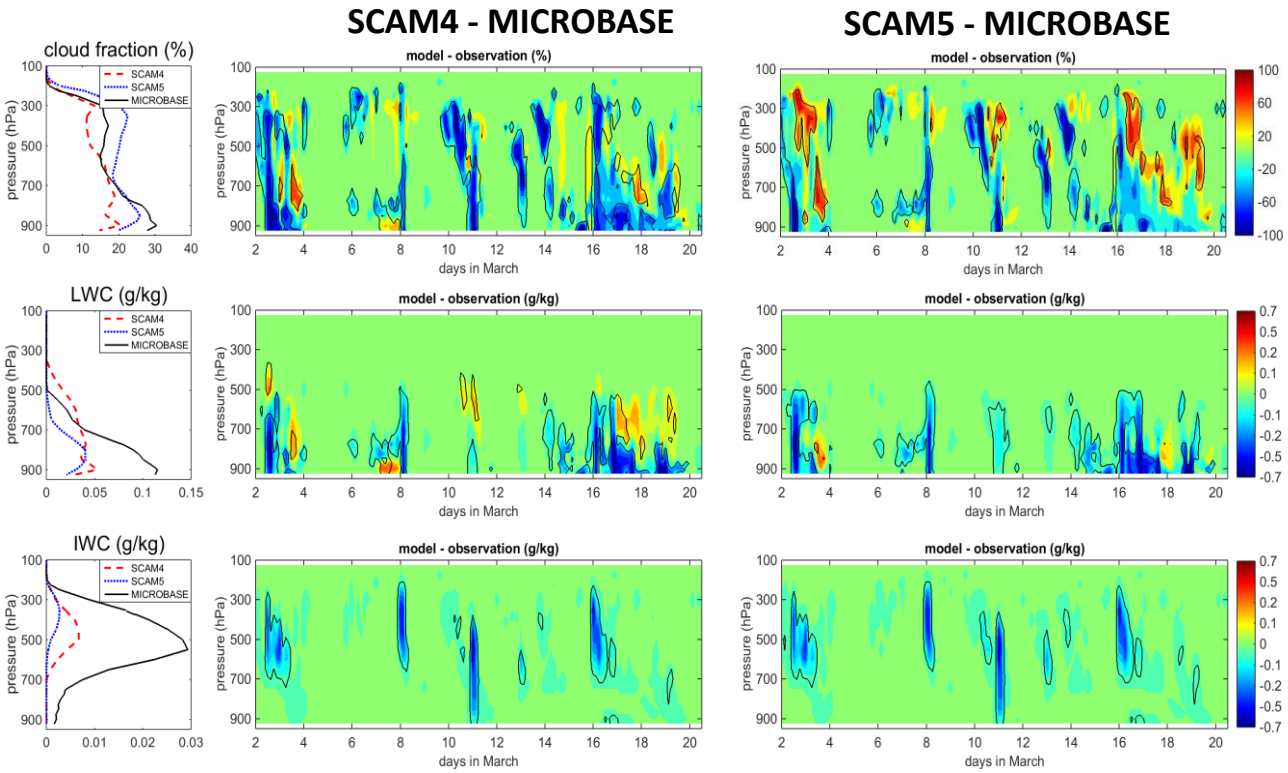

Figure 5.4: (middle and right) model bias of cloud properties at SGP site for SCAM5 (middle) and SCAM4 (right). Blue color means model underestimation while yellow/red means overestimation, with black lines circle out where model bias is larger than the total uncertainty from large-scale forcing data and observations. (left) vertical profiles averaged in March 2000 IOP. Black line is MICROBASE, red dashed line is SCAM4, and blue dotted line is SCAM5. From top to bottom: cloud fraction, LWC and IWC.

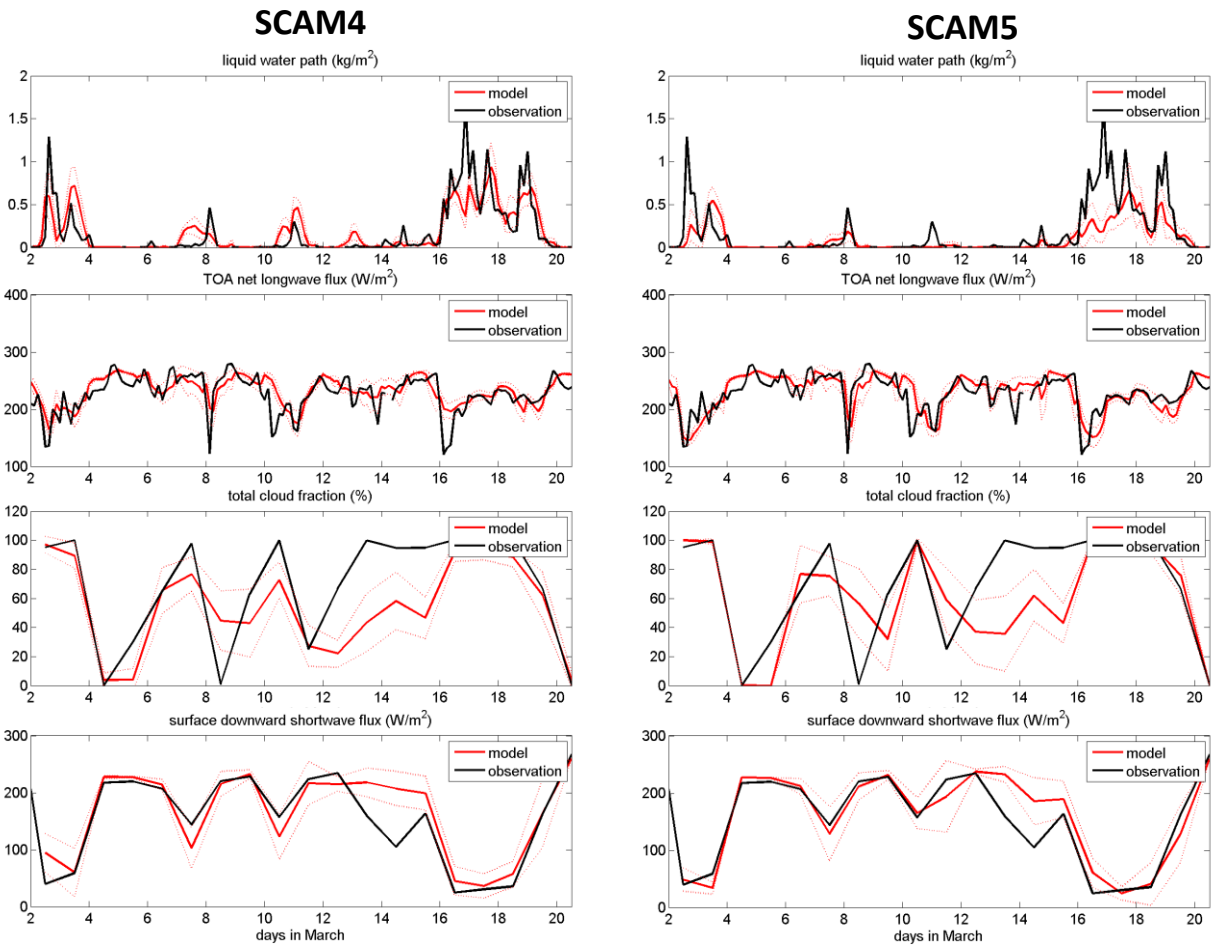

Figure 5.5: ensemble averaged simulations (red lines) of (a, b) LWP, (c, d) TOA net longwave radiative flux, (e, f) total cloud fraction and (g, h) surface downward shortwave radiative flux from SCAM5 (left) and SCAM4 (right) and observations from ARMBE (black lines). The red dashed lines show the standard deviation due to different large-scale forcing data.

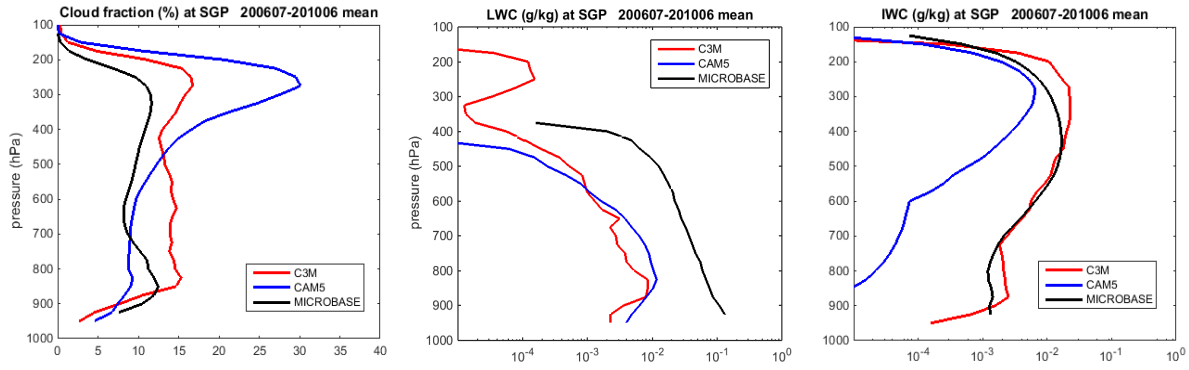

Figure 5.6: 4-year (July 2006 to June 2010) mean profiles of cloud fraction (right), LWC (middle) and IWC (right) at SGP site or the nearest grid. Blue line is from CAM5 GCM, red line is from C3M merged satellite retrieval, black line is from MICROBASE ground retrieval.

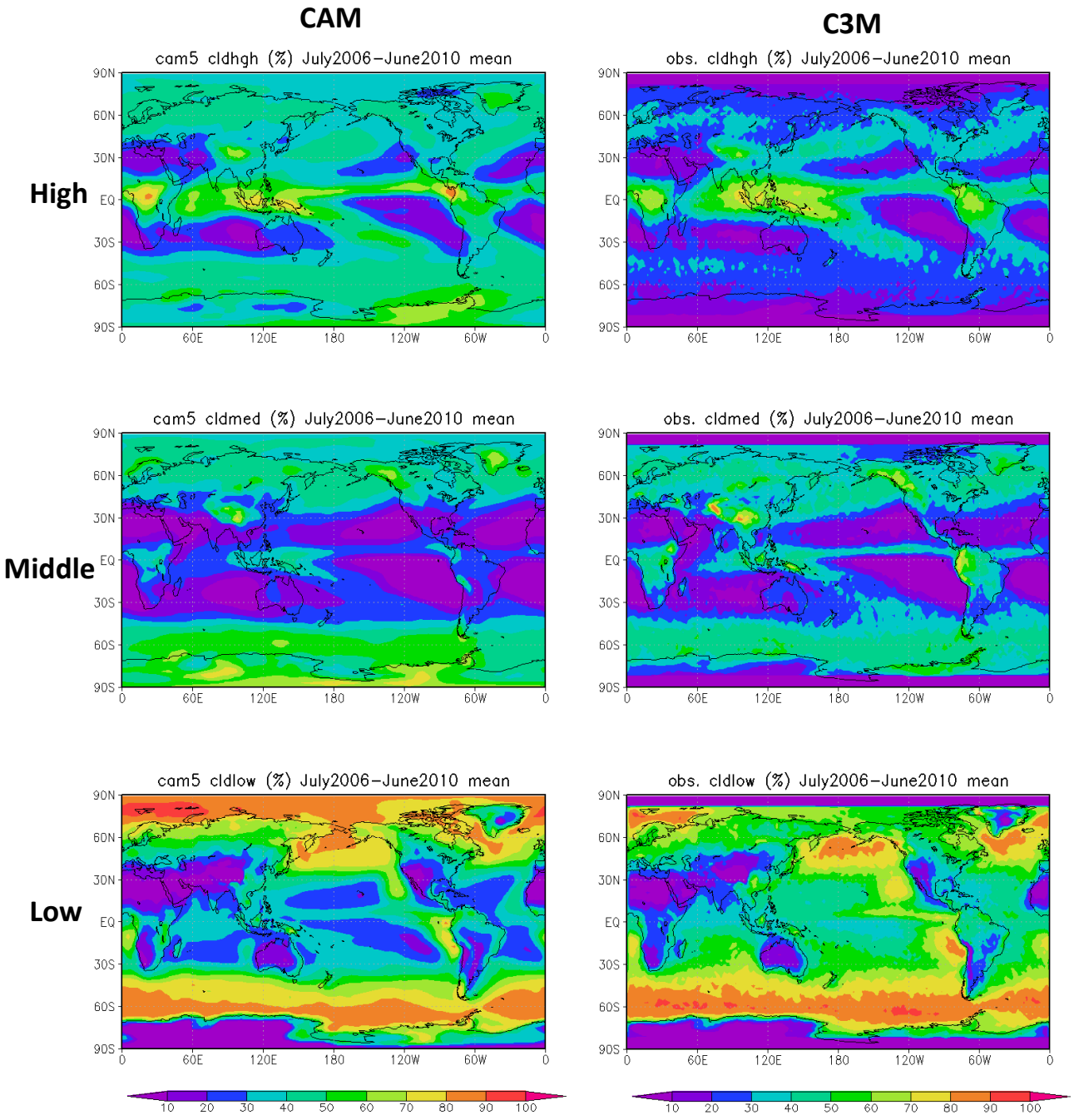

Figure 5.7: 4-year (July 2006 to June 2010) mean total high (top), middle (middle) and low (bottom) cloud fraction in CAM5 GCM (left) and C3M merged satellite retrieval (right).