ABSTRACT

Title of Document: CARBON CYCLE DATA ASSIMILATION USING A COUPLED ATMOSPHERE-VEGETATION MODEL AND THE LOCAL ENSEMBLE TRANSFORM KALMAN FILTER

Ji Sun Kang, Doctor of Philosophy, 2009

Directed By: Professor Eugenia Kalnay
Department of Atmospheric and Oceanic Science

We develop and test new methodologies to best estimate CO_2 fluxes on the Earth's surface by assimilating observations of atmospheric CO_2 concentration, using the Local Ensemble Transform Kalman Filter. We perform Observing System Simulation Experiments and assimilate simultaneously atmospheric observations and atmospheric carbon observations, but no surface fluxes of carbon. For the experiments, we modified an atmospheric general circulation model to transport atmospheric CO_2 and coupled this model with a dynamical terrestrial carbon model and a simple physical land model.

The state vector of the model prognostic variables was augmented by the diagnosed carbon fluxes CF, so that the carbon fluxes were updated by the background error covariance with other variables. We designed three types of analysis systems: a C-univariate system where CF errors are coupled only with CO_2, a

multivariate system where all the error covariances are coupled, and a one-way multivariate analysis where the wind is included in the carbon error covariance, but there is no feedback on the winds. With perfect model experiments, the one-way multivariate analysis has the best results in CO_2 analysis. For the imperfect model experiments, we applied techniques of model bias correction and adaptive inflation. With those, we obtained a high-quality analysis of surface CO_2 fluxes. Furthermore, the adaptive inflation technique also provides a good estimate of observation errors.

A new approach in the multivariate data assimilation with "variable localization", where the error correlations between unrelated variables are zeroed-out further improved the multivariate analyses surface CO_2 fluxes. We note that with the simultaneous assimilation of winds and carbon variables, we are able to transport atmospheric CO_2 with winds as well as, for the first time, couple their error covariances. As a result, the multivariate systems perform well, and do not require any kind of *a-priori* information that should be pre-calculated by independent observations or model simulations.

The many new techniques that we developed and tested put us on a solid basis to tackle the assimilation of real atmospheric and CO_2 observations, a project being carried out collaboratively by Dr. Junjie Liu under the direction of Prof. Inez Fung at UC Berkeley.

CARBON CYCLE DATA ASSIMILATION USING A COUPLED ATMOSPHERE-VEGETATION MODEL AND THE LOCAL ENSEMBLE TRANSFORM KALMAN FILTER

By

Ji Sun Kang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:
Professor Eugenia Kalnay, Chair/Advisor
Professor Ning Zeng, Co-Chair
Professor Inez Fung
Professor Brian Hunt
Professor Kayo Ide
Professor Ross Salawitch

Acknowledgements

Faculty, friends, and family members have helped me to complete this dissertation. I would like to express my gratitude to these individuals for their support and assistance.

First of all, I would like to gratefully and sincerely thank my advisor, Prof. Eugenia Kalnay, for her guidance, understanding, and friendship during my graduate studies. She supported and encouraged me to be able to enjoy research. I have learned abundant things, which is beyond description, from her. Also, my thanks go to Prof. Ning Zeng who serves as co-chair of my dissertation and provides the VEGAS and SLand models for this study. I am grateful to Prof. Kayo Ide for invaluable discussion on my work. I truly appreciate Prof. Inez Fung who helped me understand the issue on the carbon cycle at the early stage of this work. I gladly express my gratitude to Prof. Brian Hunt and Prof. Ross Salawitch for insightful discussion.

I thank to members in the 'Weather-Chaos Group', especially Dr. Takemasa Miyoshi, David Kuhl, Steven Greybush. I also want to thank Dr. Junjie Liu for providing the source code of LETKF for the meteorological fields and cheering me up during my hard times. I am thankful to Dr. Hong Li and Prof. Shu-Chih Yang for helping me make steady progress toward the current point. I am indebted to Dr. Jin-Ho Yoon especially for kindly answering my questions about a carbon cycle modeling with VEGAS. Besides, I was also greatly inspired by the students that I met in the classes of our department.

Finally, and most importantly, I want to thank my family. My parents, Nam-Hyun Kang and Hee-Ja Ahn, and my brother, Tae-Jun Kang, always trust and encourage me. I am sure that all I have done here would not have been possible without their support. I am blessed with a good family.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The atmospheric CO_2 concentration has increased from about 280 parts per million (ppm) in the beginning of industrial age to more than 380 ppm today (Figure 1.1). Since the released CO_2 in the atmosphere traps long-wave radiation emitted from the Earth's surface, the global surface temperature has increased as much as 0.6 ± 0.2°C during the last century (Houghton et al, 2001). Thus, an estimation of future atmospheric CO_2 concentration has been highlighted as essential for the projection of the future climate.

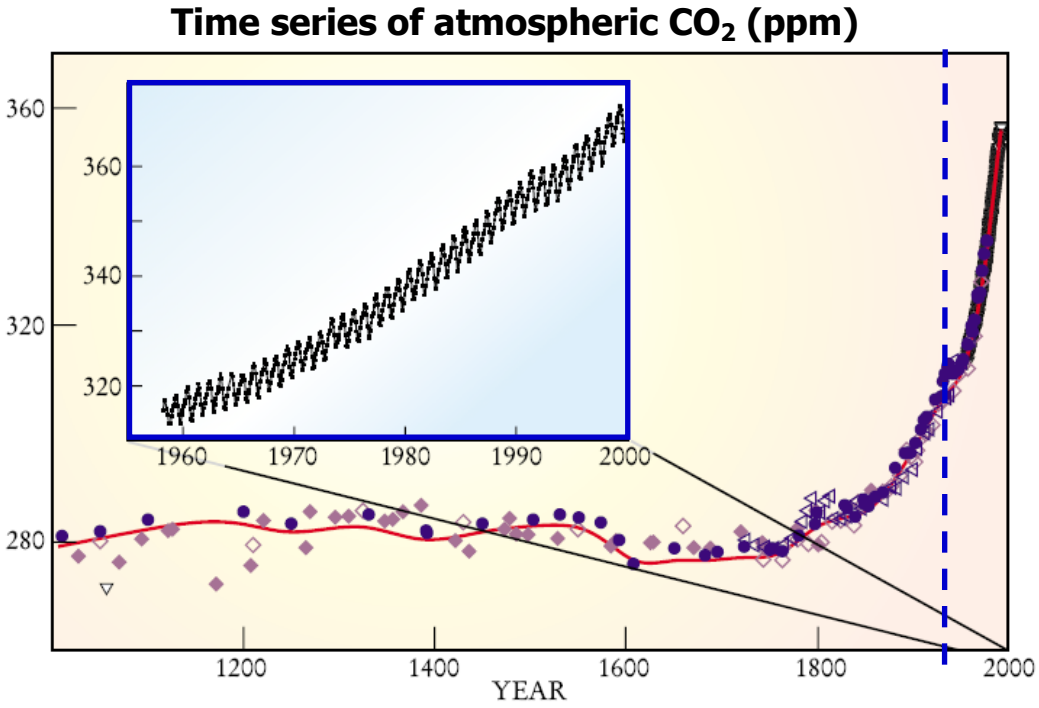

Figure 1. 1. Time Series of atmospheric CO_2 concentration: (Samiento and Gruber, 2002; Barnola, 1999, and Keeling et al., 2000)

From the comparison of growth rates between the fossil fuel emissions and atmospheric CO_2 concentration, it has been found that only about half of the emitted CO_2 remains in the atmosphere and the rest of it sequestered by the land and the

ocean (Sarmiento et al., 2002). Moreover, the increase rate of atmospheric CO_2 has a significant variability on interannual timescales (Figure 1.2). That means the capacity of the land and ocean CO_2 uptakes are varying substantially with time and are strongly connected with the climate.

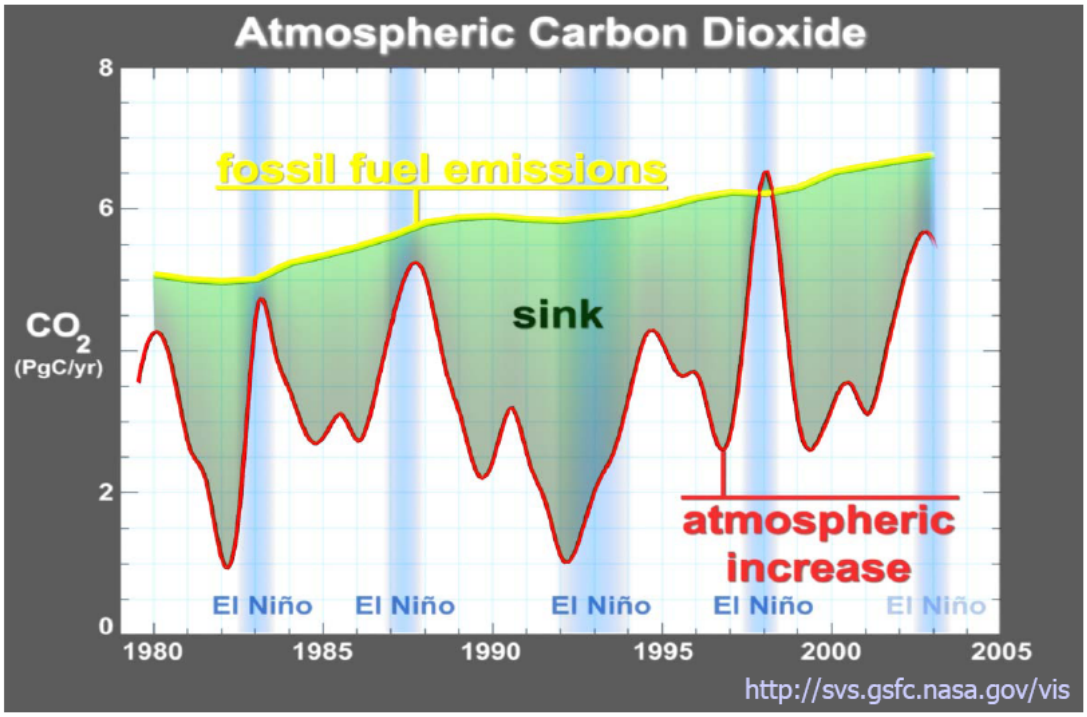

Figure 1. 2. Time series of Growth rate in atmospheric CO_2 (red) for past 25 years, and increase rate of released CO_2 by fossil fuel emission (yellow). Shaded green portion stands for the total surface CO_2 uptake and blue vertical shading indicates El Nino years.

This is a highly nonlinear problem which has complex interactions and feedbacks among all the components of land, ocean and atmosphere: the CO_2 remaining in the atmosphere is determined by the uptake of the land and the ocean, the atmospheric CO_2 concentration changes the global temperature due to its radiative properties, and the global warming caused by the increased CO_2 can have an obvious influence on the capacity of CO_2 uptake over the land and the ocean through biogeochemical processes.

So far, many studies have been done with in-situ measurements in order to understand the global carbon cycle. As one of those studies, Manning and Keeling (2006) estimated global oceanic and land biotic carbon sinks from the measurement of atmospheric O_2/N_2 ratio and CO_2 concentration over the period of 1989-2003. Observations come from the three stations of Scripps Institution of Oceanography global flask sampling network, which have the longest records in the network. The O_2/N_2 ratio and CO_2 concentration can characterize the fossil fuel combustion, terrestrial sinks and ocean uptake based on the knowledge in the chemical processes. This allows calculating how much of anthropogenic CO_2 emission is sequestered by land and by ocean quantitatively. The resultant 10-year oceanic and land biotic sinks during each period of 1990-2000 and 1993-2003 show that the ocean uptake has relatively constant value around 2.0 PgC/yr whereas the land biotic carbon sink has much greater natural variability, from 1.2 PgC/yr to 0.5 PgC/yr. These values are global total estimates.

In addition, there are studies to estimate the time-averaged CO_2 fluxes over the ocean (Mikaloff Fletcher et al., 2007, Gruber et al., 2009). As a most recent work, Gruber et al. (2009) estimates the contemporary net air-sea CO_2 flux using an inversion of interior ocean carbon observations using 10 ocean general circulation models (Mikaloff Fletcher et al., 2006, 2007). The spatial distribution of oceanic CO_2 fluxes has been estimated reasonably for 23 oceanic regions: the outgassing in the tropics, uptake in midlatitudes, and relatively small fluxes in the high latitudes even though the uncertainty in the Southern Ocean is high.

On the other hand, the terrestrial carbon uptake still remains highly uncertain in terms of its spatial and temporal variations according to the climate. However, Tans et al. (1990) has highlighted a terrestrial CO_2 uptake in the northern hemisphere to explain the north-south gradient of atmospheric CO_2 concentration. Moreover, there is much research to emphasize understanding CO_2 sequestration over the land (Bousquet et al., 2000; Gurney et al., 2002; DeFries et al., 2002; Rodenbeck et al., 2003; Friedlingstein et al., 2006) in order to figure out the interannual variability of atmospheric CO_2 (Figure 1.2) and to project the potential reservoir of CO_2 in the future.

One of the most important issues on the carbon cycle is the temporal and the spatial pattern of CO_2 sources and sinks at the Earth's surface. It is necessary to see those patterns with a finer resolution enough to understand the interaction and feedback between the climate change and the biogeochemical processes. This is an essential question in order to understand the changes in surface CO_2 fluxes under the current climate and to project the future climate.

Thus, the purpose of this study is to explore the feasibility of estimating surface CO_2 fluxes by assimilating remotely sensed atmospheric CO_2 observations using one of the advanced data assimilation methods, the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al., 2007). We investigate the analysis system to estimate surface CO_2 fluxes and atmospheric CO_2 concentration as well as atmospheric variables simultaneously on a fine temporal and spatial scale. Because this is the first test of a new methodology, this work is limited to simulated observations for an Observing System Simulation Experiment (OSSE). Only if the

results are promising in this simple approach, we will start working with real observations.

1.1. Problems in the estimation of surface CO_2 fluxes

There is a direct observation network of surface CO_2 fluxes, FLUXNET (Figure 1.3), a global collection of micro-meteorological flux measurement sites. The flux tower sites measure the exchanges of carbon dioxide between terrestrial ecosystems and the atmosphere. However, the observed fluxes are representative of areas ranging from square meters to square kilometers. Besides, there are no standard methods for aggregating flux data into various temporal scales (daily, monthly, or annual time periods) so that there are various methods used at each site for temporal scaling (http://www.fluxnet.ornl.gov/fluxnet/index.cfm). For these reasons, this dataset was unlikely to be directly used for the global analysis of surface CO_2 fluxes with an atmospheric global circulation model, although it could provide good information for the validation of the resultant analysis. Recently, Stockli et al. (2008) attempted to use this dataset to a land-surface model CLM3.5 which has been coupled with NCAR CAM3.5 model. Data from 15 sites are used to improve the model, and this study suggests that FLUXNET is a valuable tool to develop and validate land surface model.

In addition, atmospheric CO_2 concentrations have not been observed densely enough to estimate the global distribution of CO_2 sources and sinks. In situ measurements of CO_2 concentrations at ground stations have been used to monitor the carbon dioxide in the atmosphere (GLOBALVIEW- CO_2 data from ESRL/NOAA;

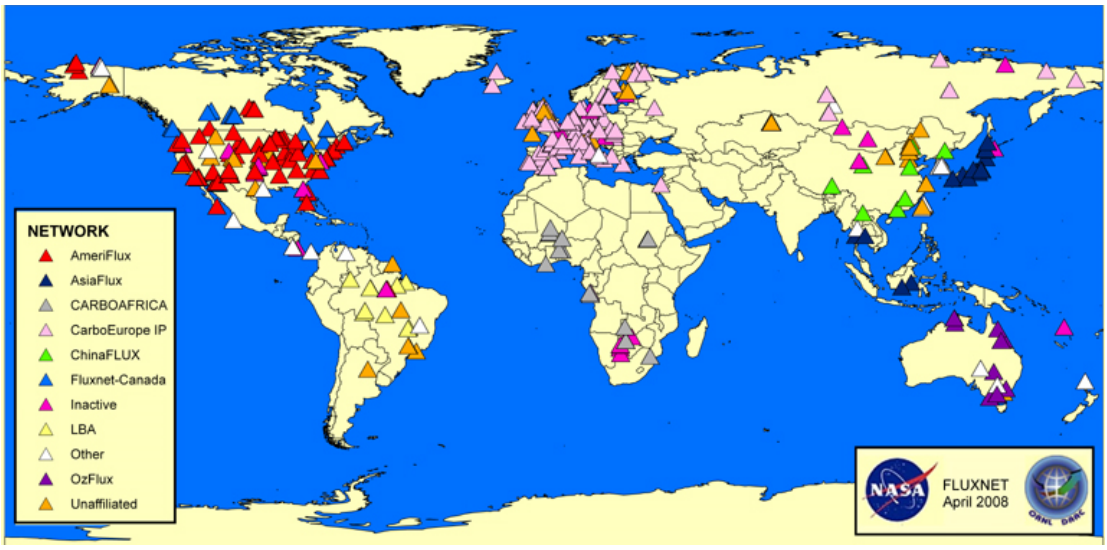

Figure 1. 3. FLUXNET network which measures the exchanges of CO_2 between terrestrial ecosystems and the atmosphere at flux tower sites (http://www. fluxnet.ornl. gov/fluxnet/index.cfm).

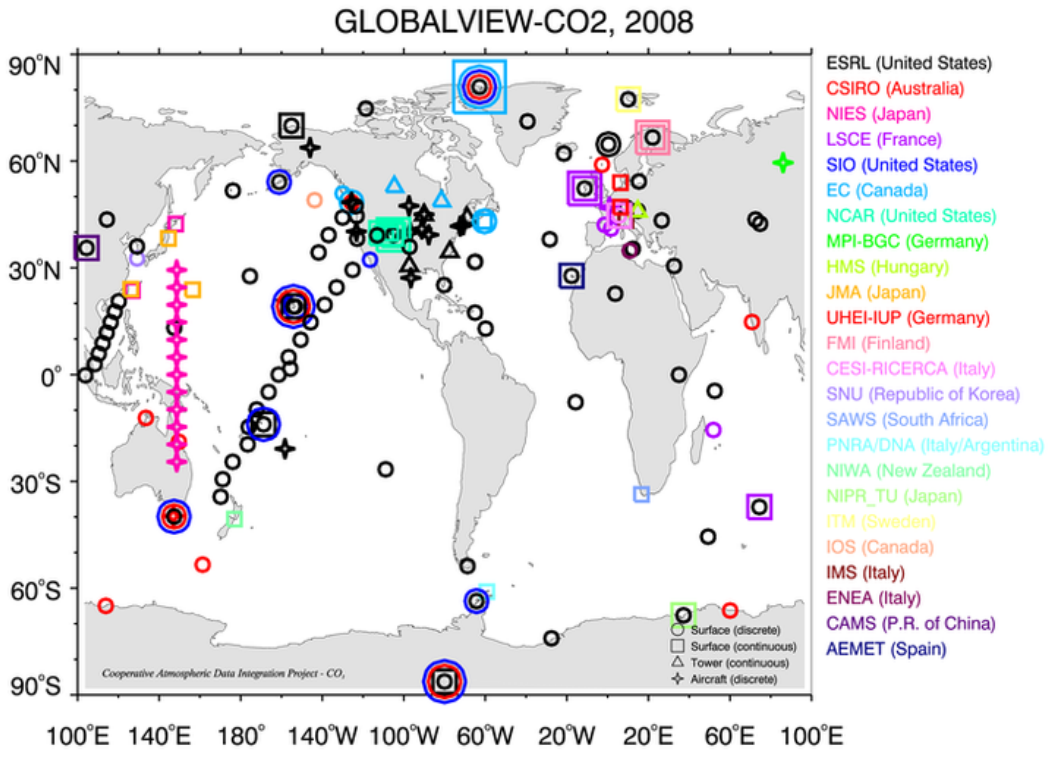

Figure 1. 4. Sampling location of GLOBALVIEW- CO_2 (http://www.esrl.noaa.gov/gmd/ccgg/globalview/co2/co2_sites.html; CO_2 concentration measurements near the surface)

Figure 1.4). The network of surface stations provides high precision information about variations of CO_2 fluxes in the global scale. This dataset has been used for a number of previous studies on the carbon cycle and hence it has contributed to extend our understanding on this field. However, the observations are spatially too sparse and temporally heterogeneous for representing the regional variations of surface CO_2 fluxes and understanding surface CO_2 sinks and sources in finer scales.

Recently, more CO_2 estimations have become available through remote sensing measurements such as the Atmospheric Infrared Sounder (AIRS; Chahine et al., 2008), the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY; Buchwitz et al., 2005), the CO_2-dedicated Orbiting Carbon Observatory (OCO; Crisp et al., 2004) and the Greenhouse gases Observing Satellite (GOSAT; Maksyutov et al., 2008, Hamazaki, 2008). Both OCO and GOSAT were designed specifically to estimate the total column of CO_2 mixing ratio which has a high sensitivity to the CO_2 near the surface. Unfortunately, the launch of OCO failed, while GOSAT was successfully launched and should start distributing data in mid 2009. These observations are expected to provide valuable information on global CO_2, much more comprehensive than the measurements (such as ground-based flask data) available so far.

In early studies on CO_2 sources and sinks, atmospheric inversion methods have been applied using an atmospheric transport model. The point-based measurement of ESRL/NOAA (GLOBALVIEW- CO_2) has been used for inferring surface CO_2 fluxes within a Bayesian framework (Gurney et al., 2004; Rödenbeck et al., 2003). The atmospheric transport model uses wind fields from a given reanalysis,

so that the impact of possible wind errors in the CO_2 transport is effectively not considered. Since the available spatial coverage of concentration data is not enough to constrain the flux estimates, *a-priori* information about the fluxes has been pre-calculated from other sources of information, such as independent measurements or model simulations. Then, the system determines the surface CO_2 fields minimizing the difference between modeled and observed concentrations and between predicted fluxes and their prior estimates. With this optimization, it is able to reduce the uncertainties in the surface CO_2 flux fields from the a-priori estimates. The spatial and the temporal resolution of the resulting estimates are still limited, due to the ill-posedness of the problem and to the limited number of available measurements (Gurney et al., 2004).

In an alternative approach to inversion methods, data assimilation techniques have recently been used to optimize the use of the observations for the analysis of surface CO_2 fluxes on the globe. Two advanced methods for data assimilation, 4-dimensional variational data assimilation (4D-Var) and Ensemble Kalman Filter (EnKF) are being used or considered for use in operational numerical weather prediction centers. Both of them have been considered for the estimation of carbon surface fluxes, and this thesis is devoted to the use of a particularly efficient EnKF method, the LETKF. As described below, the applications of data assimilation so far have been "univariate", i.e., they perform the analysis of only carbon variables and assume the winds are given from an independent analysis. By contrast, in our approach we propose to perform "multivariate" data assimilation, assimilating simultaneously both the carbon variables and the standard atmospheric variables

including winds, an approach that we will show allows estimating wind uncertainties and "errors of the day" and therefore improves the carbon variables analysis.

Peters et al. (2005) applied an ensemble Kalman filter technique (EnSRF; Whitaker and Hamill, 2002) for estimating weekly CO_2 fluxes on the surface. They assimilated ESRL/NOAA GLOBALVIEW-CO_2 data, and used the forward integration of TM5 chemistry transport model for the CO_2 forecast. The transport model was ran offline (univariately) with the meteorological fields from the European Centre for Medium Range Weather Forecast (ECMWF) model. The state vector of the analysis contains surface CO_2 fluxes at multiple time steps since observed CO_2 concentration variations contain a history of sources and sinks. Unknown surface fluxes are optimized with atmospheric observations, linked together through the atmospheric transport model, which is the observation operator within an ensemble-based data assimilation system. That is, the Kalman gain matrix determines the surface CO_2 flux fields to minimize the combination of the observation errors in the surface flask measurements and the model errors. Here, the model errors are assumed to be caused by the errors in the surface CO_2 flux fields, which force the transport model, while the system assumes a "perfect" transport. Then, the Kalman gain matrix is used to update only the mean state vector of surface CO_2 fluxes. Instead of taking account of uncertainties from the ensemble background, the covariance structure of surface CO_2 fluxes is prescribed as a 3D variational technique. Thus, their method could not take the advantage of ensemble forecast of estimating "errors of the day" with respect to the surface CO_2 fluxes. This means that their approach misses an important advantage of ensemble Kalman filter techniques. Although they update a

prior estimate from the analysis, the system was initialized by a flux estimate from the CASA biosphere model (Randerson et al., 1997) over the land and the Takahashi et al. (2002) ocean fluxes. This work was the first trial to use an ensemble based data assimilation technique for estimating surface CO_2 fluxes, and the results provided satisfactory flux estimates for the relatively large regional scales resolved by the surface flask measurement including aircraft measurements.

Another study with the Ensemble Transform Kalman filter technique (ETKF: Bishop et al., 2001) has been examined by Feng et al. (2008). This uses a chemistry transport model and scene-dependent averaging kernels of the OCO measurements (Crisp et al., 2004) as an observation operator to assimilate the 8-day mean CO_2 observation from the OCO. This is an Observing System Simulation Experiment (OSSE) so that it assimilates simulated observations. They made the state vector consist of the surface CO_2 fluxes of 144 regions on the globe at the current assimilation time step but also at the previous 11 time steps, similar to the approach taken by Peters et al. (2005). The basic concept of optimization is also like that in Peters et al. (2005), calculating the adjustment to the background based on the difference between model and observations and their uncertainties within the ensemble-based data assimilation framework. However, the analysis uncertainties have been estimated by the ensemble Transform Kalman Filter algorithm, which is different from Peters et al. (2005) that used prescribed analysis uncertainties. In this experiment, they start from *a-priori* fluxes estimated as having similar distribution but values 80% higher than the "true value", and could reduce uncertainties in the

flux estimates by 20-70% on the 144 regions overall, compared to the a priori uncertainties.

On the other hand, Baker et al. (2006) developed a four dimensional variational data assimilation approach, 4D-Var, to address the problem of surface CO_2 flux estimation, and also tested it through the use of OSSEs. They optimized time-varying boundary values of surface CO_2 fluxes over a long measurement span by the adjoint-based iterative descent method. The cost function was defined by the combination of a difference between the simulated CO_2 concentration and the observed CO_2 over the analysis window, *a-priori* errors of atmospheric CO_2 and CO_2 fluxes, and dynamic constraint between the CO_2 concentration and the flux. As a result, they could correct the errors in a priori estimate well when the observation errors and a priori flux covariance are assumed to be known perfectly, and the transport model errors are ignored. From experiments they performed to test the sensitivity to the observation error and a prior constraint on the fluxes, they concluded that the accuracy of "bottom-up" transport model is important even where there are dense observations of CO_2.

Recently, the European GEMS (Global Monitoring for Environment and Security) project has been building a comprehensive monitoring and forecasting system for atmospheric composition on both global and regional scales (Hollingsworth et al., 2008). For the data assimilation of carbon cycle in this project, it introduced a two-step process such as the atmospheric analysis first and then a flux inversion based on 4D-Var approach.

Since the observation density of atmospheric CO_2 concentration should increase with upcoming satellite data in addition to increasing surface measurements, it is not computationally feasible to use the direct inversion modeling approach that has contributed to the understanding CO_2 sources and sinks on the global scale at the earlier stage (Gurney et al. 2004; Rödenbeck et al., 2003). As the next promising methods, either ensemble Kalman filter or variational data assimilation approaches should be the technique of choice due to the computational efficiency in addition to many other advantages. So far, previous studies from both advanced data assimilation methods have used *a-priori* estimates of surface CO_2 fluxes, and none of them deals with the transport error of atmospheric CO_2 concentration forced by surface CO_2 fluxes.

For our work, we introduced a new technique for the carbon cycle data assimilation using the Local Ensemble Transform Kalman Filter (Hunt et al., 2007), in a way that does not need transport inversion or *a-priori* information. In our approach, we assimilate simultaneously all the atmospheric variables and CO_2 variables instead of using the reanalysis data of atmospheric variables when assimilating CO_2. From the background state of ensembles, we can deal with "errors of the day" and further allow the error covariance among the dynamical variables to reflect the distribution of uncertainties caused by each of variables. The details of the new method are described in Section 1.3.

This work is a part of the project on "Carbon Data Assimilation with a Coupled Ensemble Kalman Filter" supported by the Climate Change Prediction Program in Department of Energy. The objective of the project is to estimate surface

CO_2 fluxes by assimilating atmospheric CO_2 observation from space, and the research has been organized into two components: one is a simulation (OSSE) approach to develop and to test various new approaches in data assimilation using the LETKF coupled to a small primitive equations global atmospheric model, and the other is an application of methodologies tested by the simulation approach coupling the LETKF to a higher resolution Community Atmospheric Model (CAM) and with real, not simulated observations. The work of this thesis is devoted to the simulation component in this project. Thus, all the experiments here are Observing System Simulation Experiments (OSSEs). Dr. Junjie Liu is carrying out the real model, CAM3.5, real observations component, AIRS and GOSAT, of the project under the direction of Prof. Inez Fung (UC Berkeley).

1. 2. SPEEDY-C and SPEEDY-VEGAS

In the OSSEs, there should be a long model integration, known as a "nature run" assumed to be the "truth". From the nature run, we can simulate the observations which will be assimilated in the analysis. On the other hand, we need to make forecasts to create the background for the analysis. The same model can be used for both the nature run and the forecast model, or it is possible to use a different model for the forecast from that for the nature run. If we use the same model for both the nature run and the forecast, and the forecast starts from the perturbed initial conditions that are not the same as the nature run, then the departure of the forecast from the nature run can be attributed as coming only from the initial conditions. On the other hand, when we use a different model for the forecast from that for the nature

run, model errors should be considered as well in addition to the errors caused by initial condition. These experimental designs provide a very good tool to assess the performance of a new data assimilation method because in an OSSE, unlike in real life, we know the truth and we can control the errors.

In this study, we modified an intermediate-complexity atmospheric general circulation model, SPEEDY (Molteni, 2003), to simulate atmospheric CO_2 concentration with a given forcing of surface CO_2 fluxes. For the perfect model experiments, the modified SPEEDY is used for both the nature run and the ensemble forecast. Then, we coupled a dynamic terrestrial carbon model, VEGAS (Zeng et al., 2005), and a physical land surface model, SLand (Zeng et al., 2000) to the SPEEDY with atmospheric CO_2 prognostic. The coupled atmosphere-vegetation-land model is used for the nature run while the modified version of SPEEDY continues to make the ensemble forecasts in the imperfect model experiment. The forecast model does not make any changes in surface CO_2 fluxes since the surface CO_2 fluxes are the forcing term constant with time. The nature run, however, calculates the surface CO_2 fluxes every six hours through the interaction among the atmosphere, the land, and the vegetation. Thus, the changes in the surface CO_2 flux analysis only come from the data assimilation, not from the forecast model.

1.3. LETKF for carbon cycle data assimilation

1.3.1. Formulation of LETKF

LETKF (Hunt, 2005, Hunt et al., 2007) is an advanced ensemble Kalman filter data assimilation scheme. It is a square-root ensemble filter in which the observations

are assimilated simultaneously to update the ensemble mean while the ensemble perturbations are updated by transforming the forecast perturbations through a transform matrix term as in Bishop et al. (2001). The analysis is done independently at every grid point using observations from a local region, so this scheme is expected to be efficient for parallel computing systems.

$$\mathbf{y}_{[g]}^{\mathbf{b(i)}} = \mathbf{H}_{[g]}\mathbf{x}_{[g]}^{\mathbf{b(i)}} \ , i=1, \ldots , k \qquad (1.1)$$

$$\mathbf{Y}_{[g]}^{b} = \left\{\mathbf{y}_{[g]}^{b(i)} - \overline{\mathbf{y}}_{[g]}^{b}\right\} \qquad (1.2)$$

$$\mathbf{X}_{[g]}^{b} = \left\{\mathbf{x}_{[g]}^{b(i)} - \overline{\mathbf{x}}_{[g]}^{b}\right\} \qquad (1.3)$$

Here, $\mathbf{x}_{[g]}^{\mathbf{b(i)}}$ is the i-th member of ensemble forecast, $\mathbf{H}_{[g]}$ the observation operator, and $\mathbf{y}_{[g]}^{\mathbf{b(i)}}$ the i-th member of background observation ensemble. Subscript [g] indicates that the values are estimated globally and the bars above the vectors represent the mean of ensembles. Let the number of ensemble forecast be *k*, the number of observations *l*, the dimension of state vector *m*. First, the analysis system calculates the ensemble forecast on the observation locations using **H**, the global observation operator (Equation 1.1), and then computes the observation increment for every ensemble member, $\mathbf{Y}_{[g]}^{\mathbf{b}}$ (Equation 1.2). On the other hand, the deviation of each ensemble forecast from their mean is calculated (Equation 1.3). These processes are done globally initially before going to the computation for each local patch.

Now, the analysis mean state ($\overline{\mathbf{x}}^{\mathbf{a}}$) and the analysis error covariance ($\mathbf{X}^{\mathbf{a}}$) are calculated by the ensemble forecast and the observation located within each of the local patch (Equation 1.4-1.6).

$$\overline{\mathbf{x}}^{a} = \overline{\mathbf{x}}^{b} + \mathbf{X}^{b}\widetilde{\mathbf{K}}(\mathbf{y}^{o} - \overline{\mathbf{y}}^{b}) \qquad (1.4)$$

$$\tilde{\mathbf{K}} = [(\mathbf{Y}^b)^T \mathbf{R}^{-1} \mathbf{Y}^b + (k-1)\mathbf{I}]^{-1} (\mathbf{Y}^b)^T \mathbf{R}^{-1} \quad (1.5)$$

$$\mathbf{X}^a = \mathbf{X}^b [(k-1)\tilde{\mathbf{P}}^a]^{\frac{1}{2}} \quad (1.6)$$

$$\tilde{\mathbf{P}}^a = \left[(\mathbf{Y}^b)\mathbf{R}^{-1}\mathbf{Y}^b + (k-1)\mathbf{I}\right]^{-1} \quad (1.7)$$

Here, $\tilde{\mathbf{K}}$ is the Kalman gain matrix in ensemble space, and **R** is the observation error covariance, and $\tilde{\mathbf{P}}^a$ is the analysis error covariance matrix in the ensemble space. This is the system to analyze the state vector **x** which contains normally the meteorological variables such as wind, temperature, humidity and surface pressure.

1.3.2. Carbon cycle data assimilation: multivariate vs. univariate analyses

The state vector in the analysis is augmented by adding the surface CO_2 fluxes, which are then updated through the background error covariance, an approach similar to parameter estimation (Baek et al., 2006). In the formulation of LETKF, background error of surface CO_2 fluxes are not involved explicitly in calculating the Kalman gain matrix, $\tilde{\mathbf{K}}$, since they are not observed (Equation 1.5). But, background errors of surface CO_2 fluxes result in background error of atmospheric CO_2 concentration. That is, the background errors of atmospheric CO_2 concentration are partially a result of the errors in surface CO_2 forcing. Thus, $\tilde{\mathbf{K}}$ is determined in a way to minimize the errors of other dynamic variables including the atmospheric CO_2 as well as the CO_2 flux error.

According to choices of variables which are included into the state vector, $\mathbf{x}_b$, with surface CO_2 fluxes, various ways to estimate surface CO_2 fluxes are possible. When we make an analysis state vector with only the atmospheric CO_2 concentration

in addition to the surface CO_2 fluxes, the Kalman gain is calculated by only assimilating atmospheric CO_2 observations univariately. As mentioned before, the errors of surface CO_2 fluxes implicitly affect the computation of the Kalman gain matrix. Then, the surface CO_2 fluxes are updated by multiplying a background error covariance matrix of atmospheric CO_2 and surface CO_2 fluxes to the Kalman gain matrix (Equation 1.4). If the analysis state vector is designed to also include the wind fields, in addition to atmospheric CO_2 and surface CO_2 fluxes, then the analysis can reflect the background error covariance among those variables to estimate surface CO_2 fluxes multivariately. In our analysis system, atmospheric variables are assimilated simultaneously using the simulated rawinsonde observations as a version of LETKF used in Liu (2007).

In order to see how the background error covariance of the atmospheric variables with the surface CO_2 fluxes effects on the analysis of surface CO_2 fluxes, we designed various data assimilation techniques into the LETKF framework and those are introduced in Chapters 2 and 6.

1.3.3. Analysis in a presence of model error: bias correction and adaptive inflation

Since there is no model to represent the true state of atmospheric conditions perfectly, it is necessary to deal with the errors of forecast model in reality. The model bias is due to the discrepancies of a forecast model such as a coarse resolution, imperfect parameterization, etc. This study applies the low-dimensional method introduced by Danforth et al. (2007) to the bias correction of the atmospheric variables.

In practice, the ensemble forecast tends to underestimate the uncertainty in its state estimate because of model errors and nonlinearities. This leads to the underestimation of the uncertainties from the forecasts. As a result, the analysis overfits the background state estimate and gives too little weight to the observations. This inconsistency becomes larger over time, so that the information of observations is less and less used by the analysis and, eventually, it leads to an analysis that has little relationship with the observations, known as "filter divergence". Thus, it is necessary to inflate the background covariance (or the analysis covariance) during each data assimilation cycle to increase a model error covariance.

For the covariance inflation, multiplicative inflation has been applied in this work (Anderson and Anderson, 1999). It is carried out by multiplying the background perturbation from the ensemble mean by a constant factor larger than one. It is common to tune the inflation parameter manually in order to decide a reasonable value for the analysis system. But this tuning is expensive, and becomes further infeasible if the inflation factor should depend on the region or variable. Thus, Li et al. (2009) introduced a method for adaptive inflation estimation. Since the estimation of adaptive inflation is dependent on the observation errors, the paper also estimates the observation error simultaneously. The methodology presented in this paper has been found to be essential to the carbon cycle data assimilation because it is necessary to deal with the inflation for the atmospheric CO_2 separately from the meteorological variables.

Furthermore, we also consider the inflation for the surface CO_2 fluxes which do not have observations. The adaptive inflation of Li's paper is connected to the

existence of observations so it cannot be applied to estimate the inflation for the surface CO_2 flux forecast without observations. Thus, we applied a simple inflation method to surface CO_2 fluxes. It is basically to make the ensemble spread of analysis correspond to the background ensemble spread at each analysis time (similar to Zhang et al., 2004).

1.4. Outline of the thesis

Chapter 2 has the description of the model which we modified for this study and the three types of data assimilations for the carbon cycle in the LETKF framework. To test the performance of a new data assimilation system, the experiments are first done under the simple scenario given by the perfect model assumption. Next, Chapter 3 shows how a coupled atmosphere-vegetation-land model was constructed to estimate the time-varying surface CO_2 fluxes over the land. In Chapter 4, the imperfect model experiments are carried out with a method for bias correction, and an adaptive inflation and observation error estimation. The advanced adaptive inflation techniques are also applied to the perfect model simulation in Chapter 5. From the findings in Chapter 5, we introduce a new multivariate data assimilation system in Chapter 6 and see the effect of "variable localization" in the LETKF data assimilation. Chapter 7 has a summary and lessons we learned, and discusses some future research directions.

Chapter 2: Carbon Cycle Data Assimilation in the Perfect Model Simulation Using SPEEDY-C

2.1. Introduction

In reality, no forecast model is good enough to completely ignore model error, and we will have to address this serious issue, especially for the carbon cycle. But, in this chapter, we want to address the pure performance of a new analysis system for CO_2 variables with no model error or bias. To do this, we run "identical twin" Observing System Simulation Experiments (OSSEs) using a single model with CO_2. One run, called the "nature run", serves as the "truth" for the experiment. Since we will use the same model for the truth and for the forecast, there is no model error. A second run, using an ensemble data assimilation system, can then be compared to the truth. Thus, we build one forecast system for CO_2 and use it to create nature run as well as to run the ensemble forecast for a data assimilation so that it allows us to avoid the effects of model error for the moment.

In order to simulate the CO_2 concentration in the atmosphere, we modified an intermediate-complexity atmospheric general circulation model, SPEEDY (Molteni, 2003). Next, we investigated a new analysis system for the carbon cycle and tested it under the perfect model simulation to assess the performance clearly in the absence of model error.

Section 2.2 introduces the model we chose and a detailed description of modification we have done. Section 2.3 introduces the three types of data assimilation we have tested in LETKF framework. Section 2.4 describes the experimental design. The results are shown in Section 2.5. Finally, there is summary of Chapter 2 in Section 2.6.

2.2. Model: SPEEDY-C

The SPEEDY model (Molteni, 2003) is a global atmospheric, primitive equations general circulation model (AGCM). Its simplified physical parameterization schemes are computationally efficient, but maintain the basic characteristics of a state-of-the-art AGCM with complex physics. The version used for this study has triangular truncation T30 (corresponding to about 400 km horizontal resolution) with 7 sigma levels.

The original version of SPEEDY has five dynamical variables: zonal (U) and meridional (V) wind components, temperature (T), specific humidity (q), and surface pressure (Ps). To use the model for this study, we added two variables: one is atmospheric carbon dioxide (CO_2) which is treated as a tracer, so that it is affected only by the two processes of advection and diffusion, and the other is a surface flux of carbon dioxide (CF) which is a source and sink of the atmospheric carbon. Basically, CF is not changed in the model and only plays the role of forcing the atmospheric CO_2. Later, it will be updated only by the analysis step of data assimilation. Chemical processes for the atmospheric carbon dioxide have been ignored since CO_2 is one of the inert gases in the atmosphere. Moreover, there is no

feedback between the integrated CO_2 and radiative properties. Thus, the model reads the surface CO_2 fluxes as a forcing and allows it to be transported and mixed (Equation 2.1). From now on, the SPEEDY model that contains these carbon-related variables will be referred to as "SPEEDY-C".

$$\frac{\partial \mathbf{CO_2}}{\partial t} + \Im(\mathbf{CO_2}) = \mathbf{CF} \tag{2.1}$$

Equation (2.1) shows the way to calculate the tendency of atmospheric CO_2 in SPEEDY-C, where $\Im(\mathbf{CO_2})$ represents the atmospheric 3-dimensional transport and mixing, and the forcing term, **CF**, on the right-hand side of Equation (2.1) indicates the surface fluxes of CO_2. In reality, the forcing should include fossil fuel emission, land surface fluxes, ocean fluxes, and fluxes due to land use changes. In this chapter, since we are testing the ability of data assimilation to estimate surface CO_2 fluxes, we choose a very simple scenario: the source of surface CO_2 fluxes is only caused by fossil fuel emissions, which we assume to be constant in time (Andres et al., 1996).

Due to a problem with SPEEDY dynamics, based on a spectral discretization, the total amount of atmospheric CO_2 is not conserved exactly by atmospheric transports, and there is a small but significant sink of CO_2 concentrated in the Southern Hemisphere stratosphere. Since lack of conservation is a well known generic problem, especially for spectral models, and it is desirable to conserve total CO_2, we opted for making a simple correction. After the model reads the surface CO_2 flux fields, they are converted to the atmospheric CO_2 and then transported and mixed in the atmosphere. Thus, we could calculate how much the total amount of atmospheric CO_2 should be with the given forcing of surface CO_2 flux fields. We also computed the actual amount of global atmospheric CO_2 in the SPEEDY-C

simulation. From this, we could estimate the ratio of the total amount of atmospheric CO_2 which the model should have to what the model actually has. By multiplying the atmospheric CO_2 by this ratio at every grid point and every time step, we can get an increase of simulated atmospheric CO_2 concentration with a given forcing. Although this is not an ideal correction, it maintains conservation of total CO_2, but with a small global redistribution.

2.3. Three types of data assimilation techniques

So far, estimations of surface fluxes of carbon have been made univariately, with inversion methods or with data assimilation systems that assume that the wind is given by a reanalysis, and do not couple CO_2 and wind errors (Chapter 1). Here we will compare such univariate approach with a multivariate approach in which the estimated errors of CO_2 are coupled with the other atmospheric estimated errors within the background error covariance. As far as we know, this is the first time that this has been done, even in simulation mode. Since the CO_2 errors estimated in the EnKF data assimilation may have large sampling errors, we found it desirable to create a "one-way" multivariate system in which CO_2 errors do not provide feedback to the winds.

2.3.1. Carbon-univariate data assimilation

For the **carbon-univariate (C-univariate)** data assimilation, atmospheric CO_2 concentration and surface CO_2 fluxes are updated only by these two variables, and are not affected by other atmospheric variables. That is, there are two separate

analysis systems: one for the atmospheric variables, and the other for the CO_2-related variables. These two systems never talk to each other during the analysis (similar to the univariate CO_2 assimilation that assumes winds are given by another reanalysis). The system for the atmospheric variables has dynamic variables (state vector) $\mathbf{x}_1 = \mathbf{(U, V, T, q, Ps)}$, while the one for the CO_2 variables has $\mathbf{x}_2 = \mathbf{(CO_2, CF)}$ as a state vector in the analysis cycle (Equation 1.1-1.6.). Figure 2.1 is a schematic plot to show the background error covariance matrices used for those analyses. Diagonal components of those matrices indicate the error variance of each variable while the off-diagonal components are the correlation between the variables. Black boxes indicate that there is no correlation allowed between the variables. From this, the pink box of Figure 2.1(a) allows only the background error covariance between atmospheric CO_2 and surface CO_2 fluxes to produce their analyses while the errors of all atmospheric variables are coupled in a green box of the plot. As indicated above, this approach is similar to the "carbon-univariate" approaches that have been used so far to perform carbon data assimilation (Peters et al., 2005; Baker et al., 2006; Feng et al., 2008) or inversions (Bousquet et al. 2000; Gurney et al. 2004; R□denbeck et al., 2003).

2.3.2. One-way multivariate data assimilation

Next, we consider a **one-way multivariate** data assimilation in which the atmospheric CO_2 concentration and surface CO_2 fluxes are updated by these two carbon variables as well as the wind fields, while the wind field in addition to other atmospheric variables such as temperature, specific humidity and surface pressure is

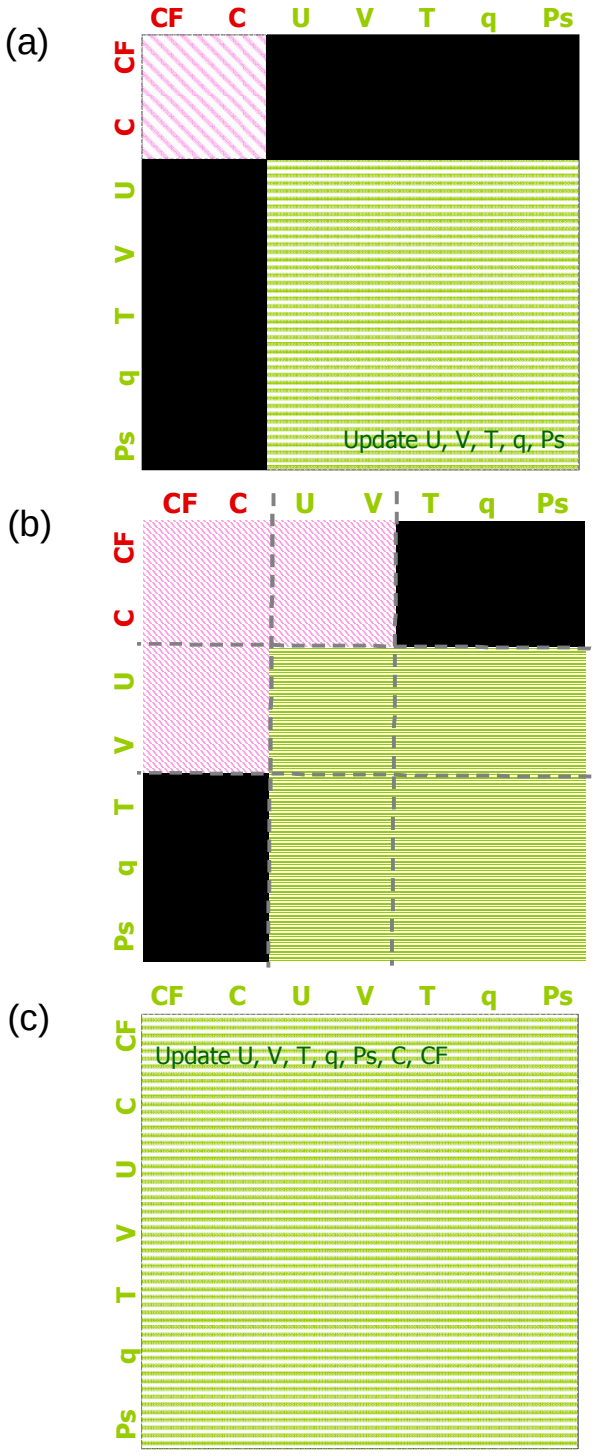

Figure 2. 1. Schematic plots of the background error covariance matrix for (a) C-univariate, (b) 1-way multivariate, (c) multivariate data assimilations. (C: atmospheric CO_2, CF: surface CO_2 fluxes)

not affected by these two carbon-related variables. As indicated above, this is done to minimize spurious feedback due to sampling errors in the estimation of CO_2 errors.

This method also has two analysis systems, and the system for the atmospheric variables is exactly same as one in the C-univariate data assimilation. For the CO_2 variables, however, we made the state vector of $\mathbf{x_2} = \mathbf{(U,V,CO_2,CF)}$ to allow the flow-dependent errors estimated for the wind fields to provide feedback to the CO_2 variables. However, the wind field from the analysis with $\mathbf{x_2}$ is discarded and we update only CO_2 and CF from the system of $\mathbf{x_2}$. That is, the pink box of Figure 2.1(b) includes the background error of wind but we only save the analyses of CO_2 and CF from the pink box. The wind fields are updated by the green box as in the C-univariate analysis. Thus, information from the wind field is given to the carbon-related variables but the information from CO_2 variables does not cause any change in the analysis of the wind field as well as other atmospheric variables. This method was designed because the atmospheric CO_2 is transported and diffused by the wind field, but is not influenced by the other atmospheric variables in the forecast model, and at the same time it prevents sampling errors in the CO_2 estimation from contaminating the winds.

2.3.3. Multivariate data assimilation

In this method (which is the standard approach that would be taken in EnKF systems), all the dynamical variables are included in one vector $\mathbf{x}$ so that the analysis of every variable is determined by the background error covariance among all variables. In other words, there is only one analysis system and the state vector

is $\mathbf{x} = \mathbf{(U, V, T, q, Ps, CO_2, CF)}$. Thus, the background error covariance matrix has the shape of Figure 2.1(c).

With this methodology, we allow the atmospheric CO_2 to be analyzed by the updated background error of wind field simultaneously, not using reanalysis winds like most of the other previous studies. Furthermore, we can assimilate simultaneously all the atmospheric variables and CO_2 variables. This is because we do not use any inversion to calculate the back trajectory of atmospheric CO_2 concentration in order to estimate surface CO_2 fluxes using the wind fields. Most of research on this issue uses the inversion method (Enting, 2002) which requires the wind fields, mainly from a reanalysis, in order to estimate surface CO_2 fluxes, whereas we do assimilate atmospheric CO_2 concentration and other meteorological variables simultaneously so that we calculate the Kalman gain matrix with a background error of all the dynamic variables including surface CO_2 fluxes.

2.4. Experimental Design

Under the perfect model assumption, the SPEEDY-C was used to create a "nature" (truth) run. We create observations from this nature run by adding random perturbations. At the same time, SPEEDY-C with six prognostic variables including atmospheric CO_2 concentration is used as the forecast model in which the CF over land is updated only by the analysis. Again, the forecast model does not have a dynamical forecast equation for CF, so the forecast of CF is persistence (starting from

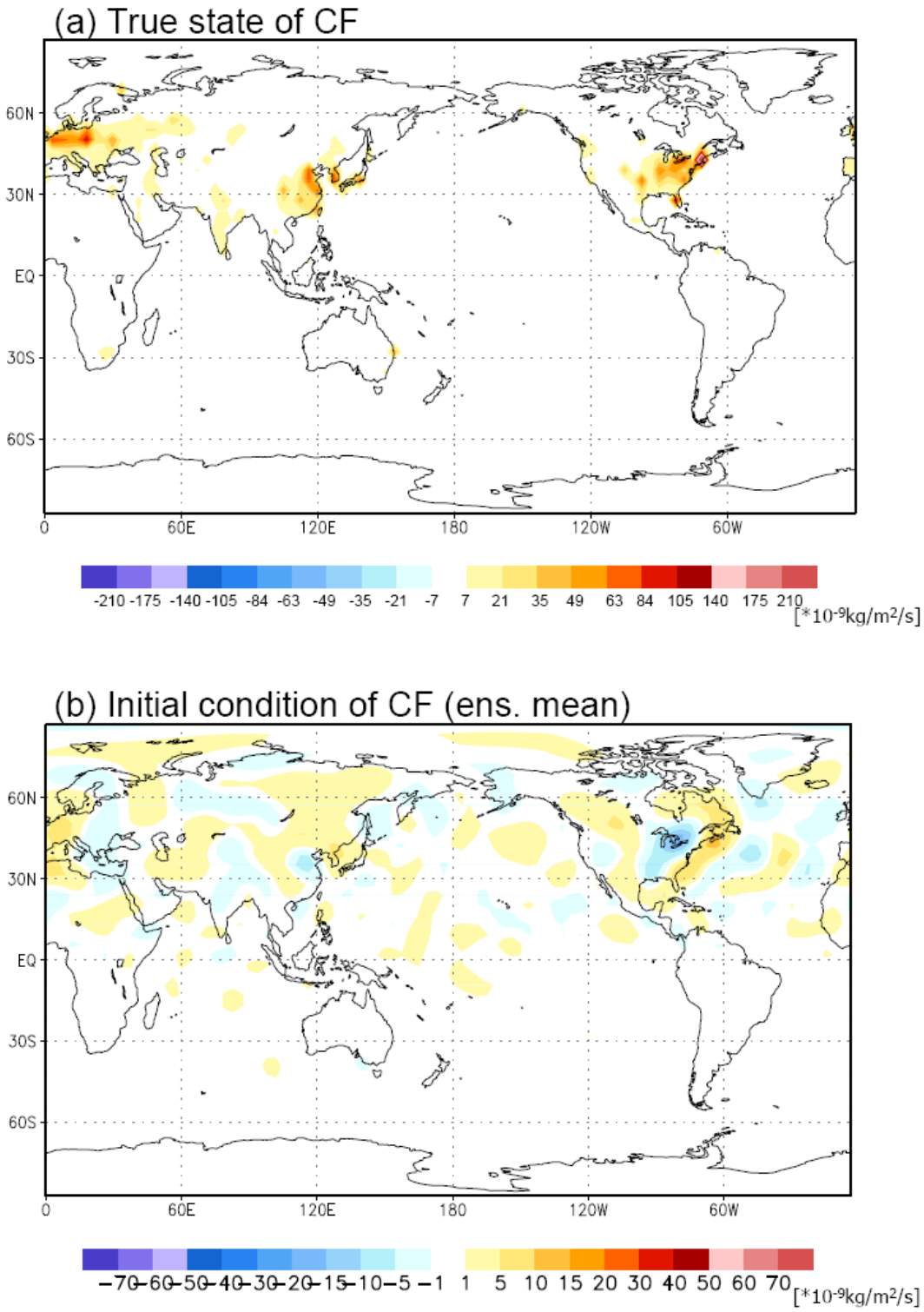

Figure 2. 2. (a) A true state of surface CO_2 fluxes which includes only anthropogenic emission as a constant forcing with time, (b) Initial condition of surface CO_2 fluxes

the analysis value). All the experiments here only include fossil fuel emission (Andres et al., 1996) with a total of 6 PgC/yr as a constant forcing with time. The spatial distribution of this forcing is shown in Figure 2.2(a). This is what we want to estimate through the analysis.

The initial conditions of (U, V, T, q, Ps, CO_2) for the 20-ensemble forecast are created by adding random perturbations to a state in the truth run which were chosen randomly in time. The standard deviation of the random perturbations used for the initialization depend on the scale of each variable: 1 m/s for U and V, 1 K for T, 0.1 g/kg for q, 1 hPa for Ps, and 1.0 ppmv for CO_2. The initial condition of the surface CO_2 fluxes has been generated separately as follows: from 20 fields of CO_2 concentration in the lowest three layers at arbitrary times, we subtract the one-day prior state of CO_2 concentration, and then convert the units of the field from the ppmv/day to the $kg/m^2/s$ (Figure 2.2). This approach was found necessary because initializing the fluxes with random numbers (as we first attempted to do) failed to converge to satisfactory results. As suggested by Zupanski et al., 2006, when the spatial scale of initial perturbation is too small to represent physically meaningful signals, EnKF can result in erroneous solutions. To avoid underestimating the uncertainty in the ensemble space, a constant 5% multiplicative inflation was applied. This value had been previously found to be optimal for the atmospheric variables data assimilation (Liu, 2007). The assimilation cycle time is every six hours.

The observations for all the experiments were simulated by adding random perturbations to the "nature" run, with the same magnitude of random perturbations as those used for the initial conditions. For the atmospheric variables, observations

have the spatial distribution of the rawinsonde network, where coverage of grid points is about 9 % globally (Figure 2.3). Atmospheric CO_2 concentration is observed at every other grid so that the coverage is about 25% in the horizontal, an optimistic assumption because we wanted to explore the potential of EnKF first in a favorable scenario. According to the vertical resolution of CO_2 observations, we performed three kinds of experiments: "ALL LEVELS", "OCO + AIRS", and "OCO" experiments with different vertical resolusions.

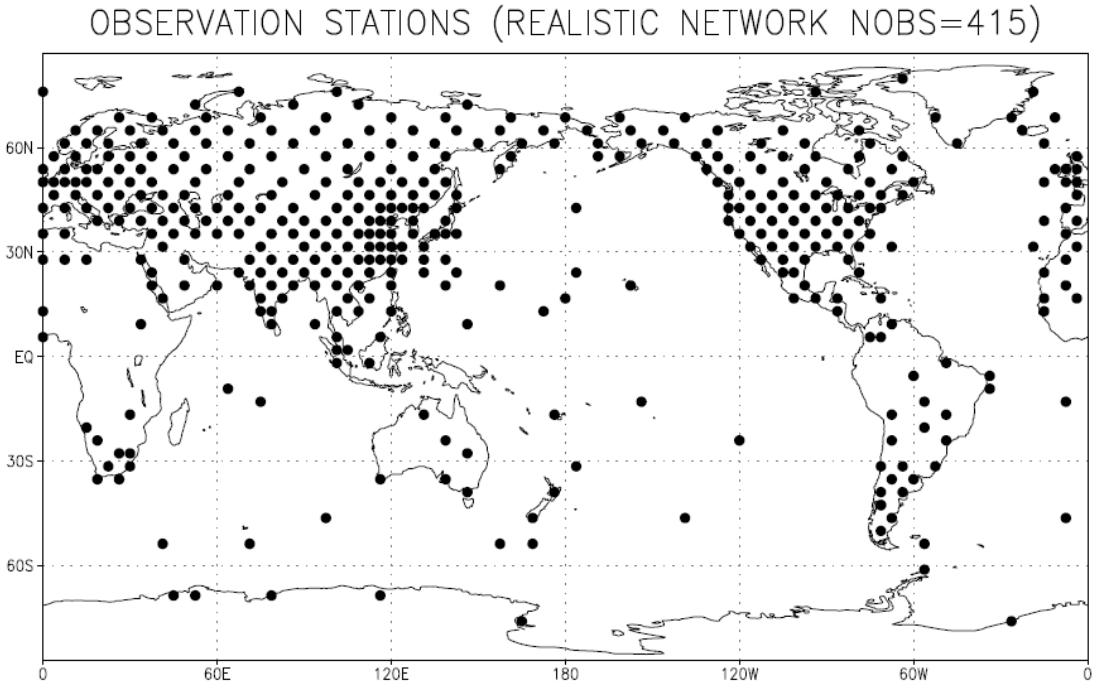

Figure 2. 3. Distribution of rawinsonde observation network.

2.4.1. "ALL LEVELS" experiment

At first, the atmospheric CO_2 concentration was assumed to be observed at every vertical level. For this case, all three data assimilation approaches introduced in Section 2.3 were tested.

2.4.2. “OCO + AIRS” experiment

Since the OCO instrument was known to be most sensitive to the CO_2 concentration in the lower troposphere (Crisp et al., 2004) and CO_2 retrieval from AIRS (Maddy et al., 2008) has instead the largest sensitivity in the middle troposphere near 7~9 km in the vertical (Figure 2.4), the experiments in this section were designed to have the CO_2 observation at only two layers, the lowest layer (σ=0.95) and at the mid-troposphere (σ=0.34). For this case, only the one-way multivariate data assimilation was performed.

These experiments were performed before the OCO launch failed, but since we will still have access to GOSAT measurements (Hamazaki et al., 2008) and other prospective satellite measurements which have a sensor most sensitive to atmospheric CO_2 near the surface, this experimental set-up is still important.

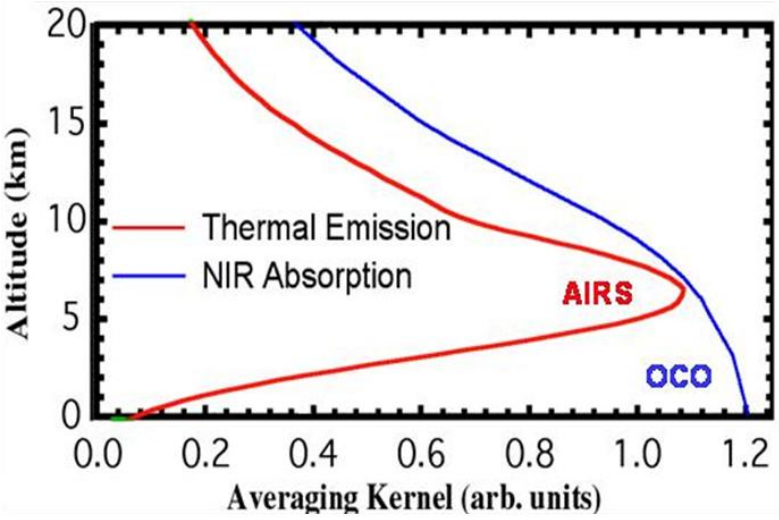

Figure 2. 4. Representative vertical averaging kernels for column CO_2 soundings using near IR absorption of reflected sunlight in the 1.61-lm CO_2 band (blue) and thermal IR emission near 14.3 lm (red). Thermal IR soundings are less sensitive to near-surface CO_2 because of the small surface–atmosphere temperature contrast. (Crisp et al., 2004)

2.4.3. “OCO” experiment

In order to see how important it is to have good measurements near the surface for the analysis of surface CO_2 fluxes, we only assimilate the CO_2 observations on the lowest layer where OCO was known to be most sensitive. For this, only the one-way multivariate data assimilation was done.

2.5 Results

We want to point out that the initial condition of surface CO_2 fluxes used here does not include any *a-priori* information. Figure 2.2 shows that the initial surface CO_2 fluxes have not only a different spatial pattern but also an inconsistent magnitude compared to the true state. This is in contrast from other previous studies which require a reasonable initial estimation for surface CO_2 fluxes.

2.5.1. Performance of SPEEDY-C

In order to see whether SPEEDY-C simulates atmospheric CO_2 reasonably well, we made a comparison with the results of an experiment made with NCAR CCM (Community Climate Model) provided by Dr. Fung. First, one needs to point out that NCAR CCM is a much more sophisticated model than SPEEDY-C in terms of physics and dynamics, and its resolution is also higher. CCM has a spectral resolution of T42 (2.8°×2.8°) in the horizontal and 18 layers in the vertical. Considering these differences, the results shown in Figure 2.5 suggest that carbon simulations with SPEEDY-C are sufficiently realistic for this study.

Figure 2.5 shows the result of an experiment which has only fossil fuel emission with 6 PgC/yr and has been started from zero state of atmospheric CO_2. The top panels are an annual mean of atmospheric CO_2 over the third year on the surface layer from SPEEDY-C (left) and NCAR CCM (right), and bottom panels are a vertical cross section of zonal mean CO_2 concentration at the beginning of third year (left: SPEEDY-C, right: NCAR CCM). From this figure, we can see the spatial distribution of CO_2 simulated by SPEEDY-C generally agrees with that of NCAR CCM even though the mixing tends to be stronger in SPEEDY-C with a deeper surface layer. Also, SPEEDY-C represents the well-mixed CO_2 within deeper surface

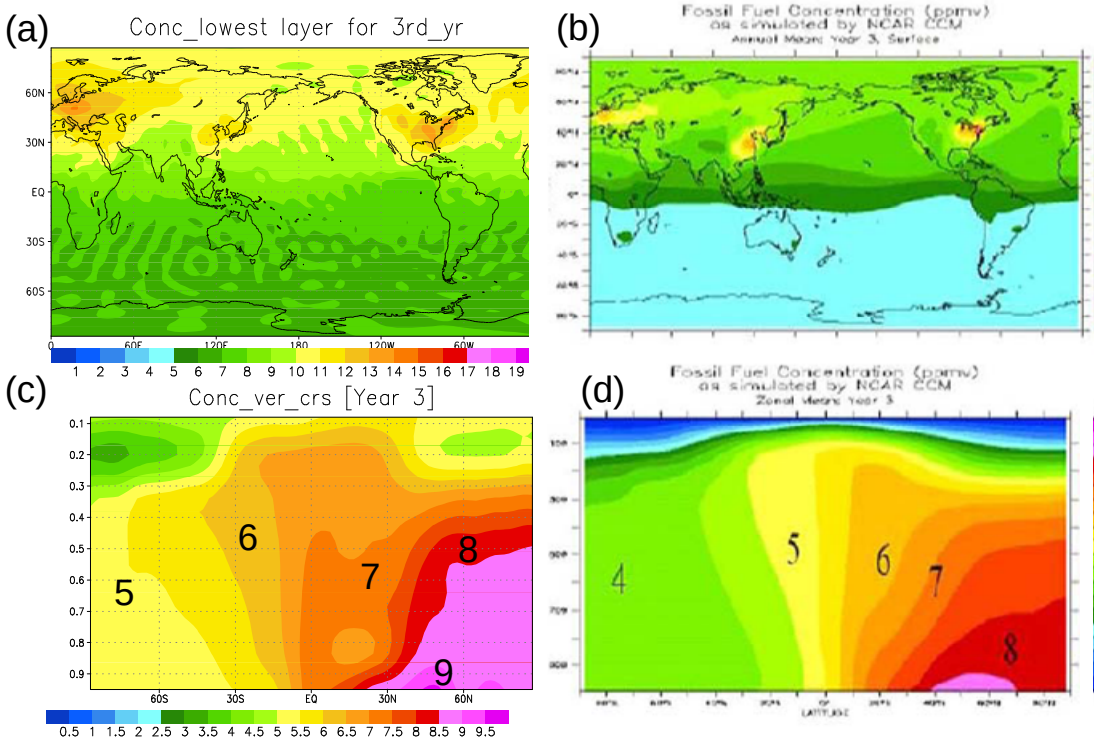

Figure 2. 5. Comparison of SPEEDY-C with NCAR CCM: annual mean of atmospheric CO_2 concentration in the surface layer for third year in (a) SPEEDY-C, and (b) NCAR CCM, and the vertical cross section of zonal mean at the beginning of 3rd year simulation in (c) SPEEDY-C, and (d) NCAR CCM. The experiment starts from zero state of CO_2 in the atmosphere with a fossil fuel emission of 6 PgC/yr. (unit: ppmv)

layer. Also, SPEEDY-C represents the well-mixed CO_2 within a hemisphere and a large gradient between hemispheres as in NCAR CCM. As a general understanding of CO_2 transport in the atmosphere, we also confirmed that CO_2 mixing within a hemisphere takes about three months and it takes about a year to have significant CO_2 mixing between hemispheres.

2.5.2. Analysis of CO_2 variables with LETKF

2.5.2.1. "ALL LEVELS" experiment

The three types of data assimilation introduced in Section 2.3 are examined in the ALL LEVELS experiments to allow comparing the performance of each data assimilation scheme. Figure 2.6 shows the global RMS error of the dynamic variables from three of the analyses. The standard atmospheric variables are analyzed similarly well with the rawinsonde distribution of observations through the three data assimilation methods, while the results of CO_2 variables vary for each of schemes. As we expected, the results of the atmospheric variables from the C-univariate data assimilation are exactly the same as those from the one-way multivariate data assimilation.

For the carbon-related variables, the carbon-univariate data assimilation fails to analyze both CO_2 and CF and we have "filter divergence". Since the C-univariate data assimilation for the CO_2 variables has only these two variables in the background error covariance matrix while there is no observation for CF, the system seems to suffer from the lack of information so that the analysis of atmospheric CO_2

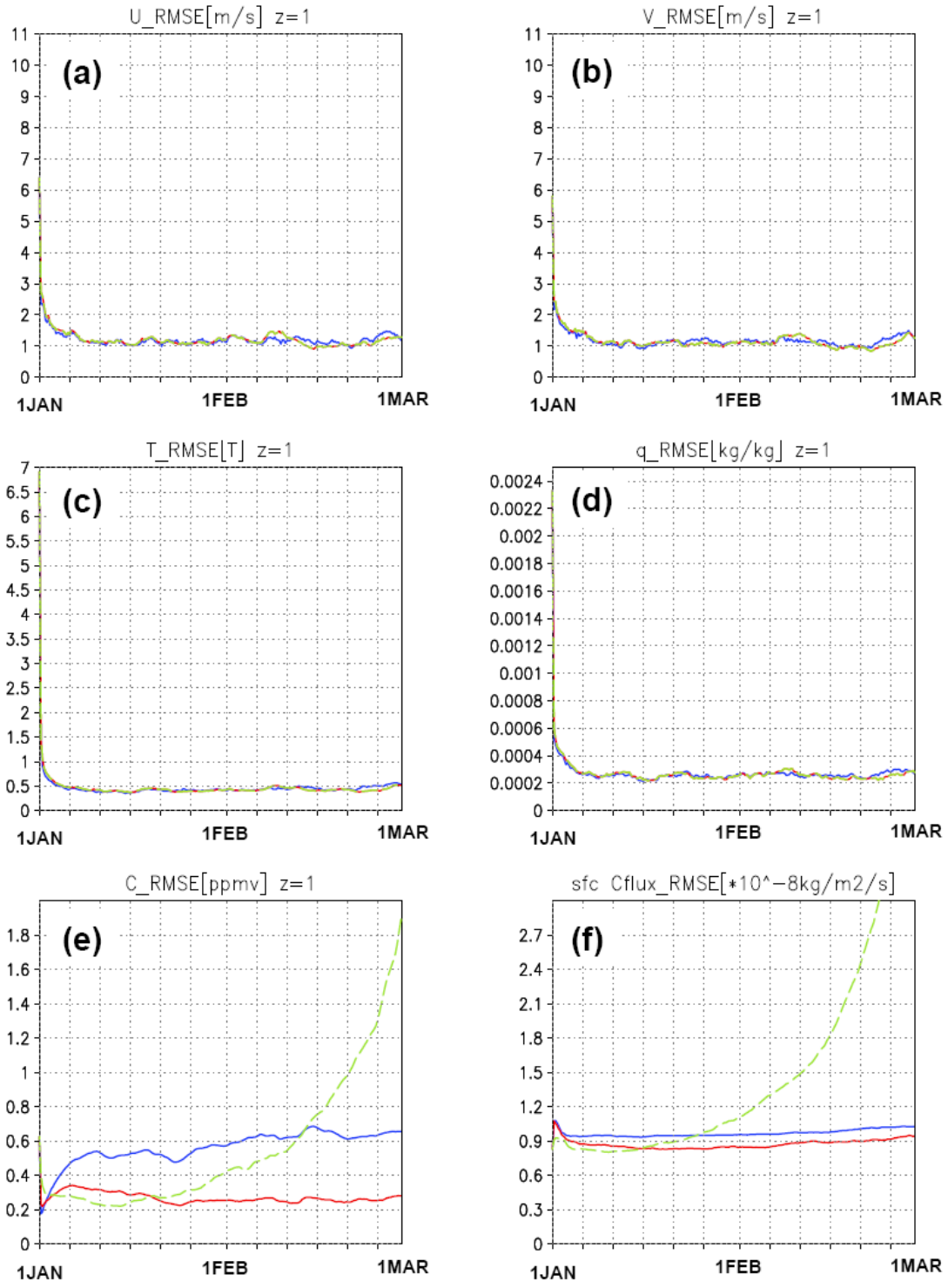

Figure 2. 6. RMS error of analysis from three types of data assimilation: uncoupled (green), multivariate (blue), and one-way multivariate (red) data assimilation for (a) U (m/s), (b) V (m/s), (c) T (K), (d) q (kg/kg), (e) atmospheric CO_2 (ppmv) on the lowest layer of model, and (f) surface CO_2 fluxes (10^{-8} kg/m^2/s).

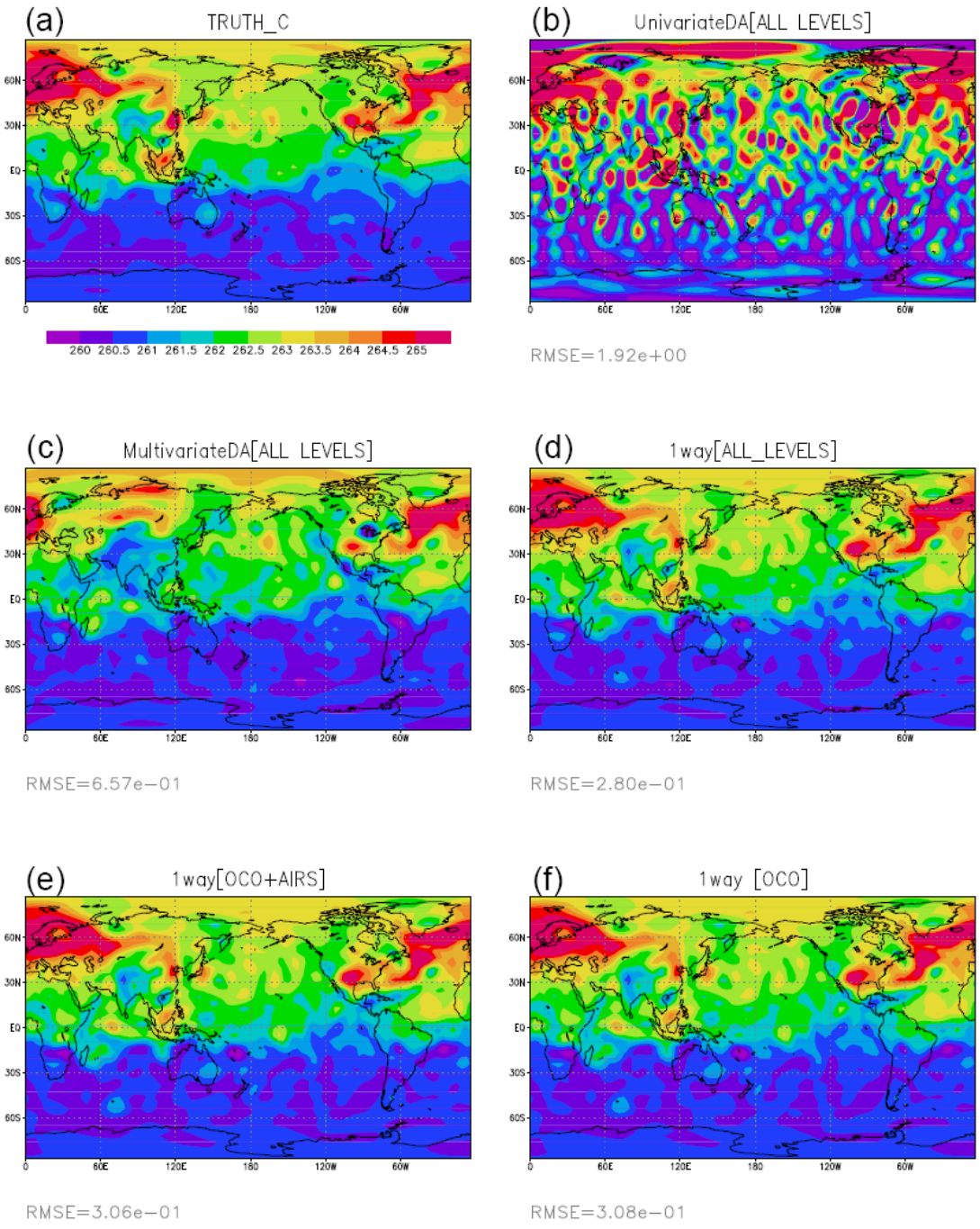

Figure 2. 7. (a) True state of atmospheric CO_2 on the lowest layer of model after two months from the start of analysis, and the resultant analysis of it from (b) uncoupled data assimilation under ALL LEVELS experiment, (c) multivariate data assimilation under ALL LEVELS experiment, (d) one-way multivariate data assimilation under ALL LEVELS experiment, (e) one-way multivariate data assimilation under OCO+AIRS experiment, and (f) one-way multivariate data assimilation under OCO experiment. The number in the left –bottom of each figure is RMS error [unit: ppmv]

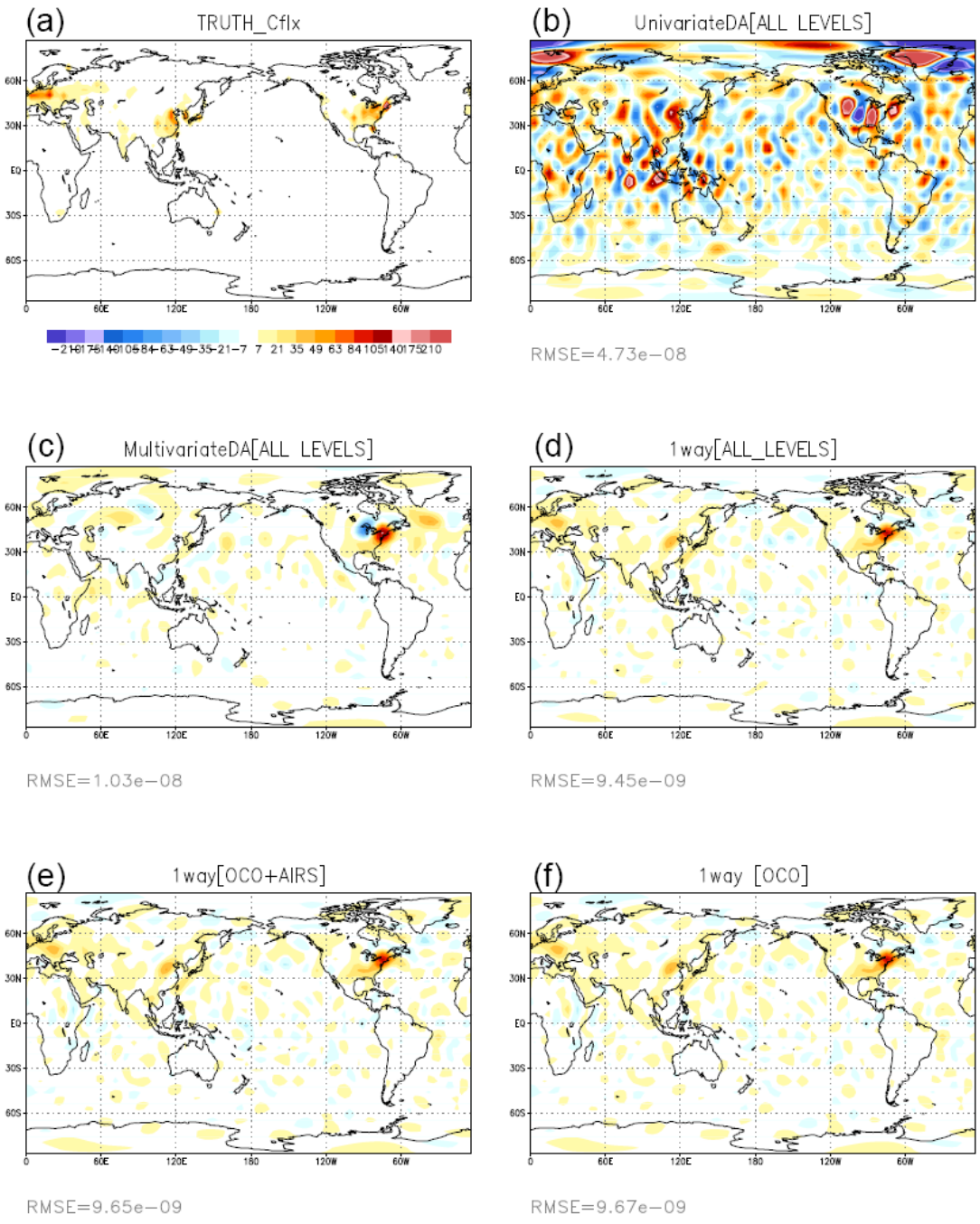

Figure 2. 8. (a) True state of surface CO_2 fluxes after two months from the start of analysis, and the resultant analysis of it from (b) uncoupled data assimilation under ALL LEVELS experiment, (c) multivariate data assimilation under ALL LEVELS experiment, (d) one-way multivariate data assimilation under ALL LEVELS experiment, (e) one-way multivariate data assimilation under OCO+AIRS experiment, and (f) one-way multivariate data assimilation under OCO experiment. The number in the left –bottom of each figure is RMS error [unit: 10^{-9}kg/m^2/s]

concentration performs poorly even with observations at every other grid point. That is, the carbon-related system of CO_2 and CF does not have enough constraints so that a bad analysis of one variable can cause negative feedback to the other variable when one of them goes wrong (for example, as shown in Figures 2.7(b) and 2.8(b), when the errors in CO_2 fluxes overwhelm the analysis). Moreover, the current experiments use a fixed inflation factor with time so the analysis system cannot use the observation information flexibly. For these reasons, the analyses of both CO_2 variables diverge at the end. We tried to increase inflation for the C-univariate data assimilation in order to stabilize the filter, but the analyses of CO_2 variables actually diverged even faster with larger inflation factors (Figure 2.9). That is because the large inflation factor gives less contribution of forecast to the analysis while the observation is insufficient to constrain CO_2 variables. Thus, the analysis system for CO_2 variables has difficulties in combining information from both forecast and observation with a large inflation factor.

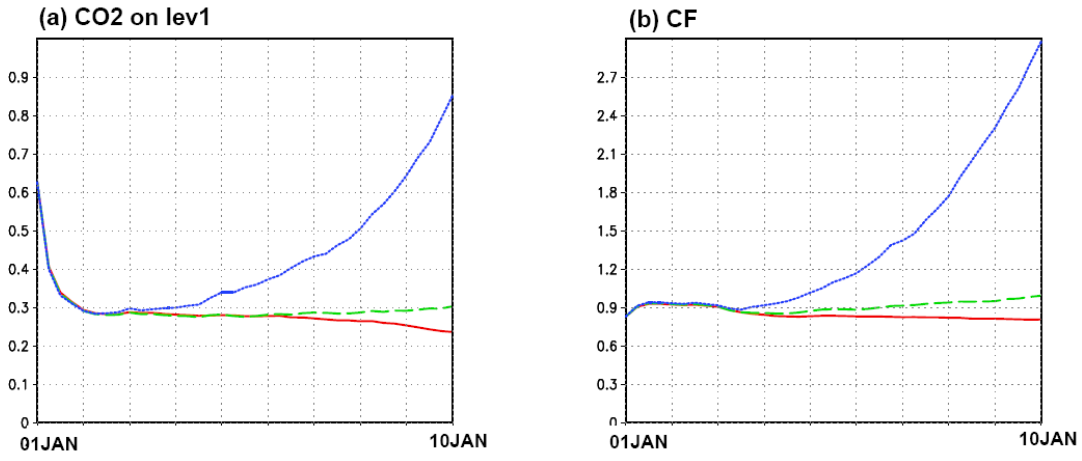

Figure 2. 9. RMS errors for the first 10 days of (a) CO_2 concentration on the lowest layer and (b) surface CO_2 fluxes from the C-univariate (uncoupled) data assimilation with 5% (red: control), 15% (green), and 30% (blue) of multiplicative inflation.

On the other hand, the multivariate data assimilation has better performance than the C-univariate one in terms of both RMS error (Figure 2.6) and spatial distribution (Figures 2.7 and 2.8). Since the errors of all variables are coupled, the analyses of CO_2 variables can have more constraint from the observation of other variables. Thus, the analyses of CO_2 variables converge to the true state unless the estimate of surface CO_2 fluxes gets far from the true state. Indeed, we have more stable and better performance on both atmospheric CO_2 concentration and surface CO_2 fluxes from the multivariate data assimilation than from the uncoupled one. The analysis of surface CO_2 fluxes, however, has rather large spurious negative values over the eastern US (Figure 2.8(c)) and these fluxes result in the degradation of the atmospheric CO_2 analysis (Figure 2.7(c)).

The performance of one-way multivariate data assimilation is optimal for the CO_2 variables (Figure 2.6). The negative values of surface CO_2 fluxes shown in the multivariate data assimilation disappear in this scheme so that the analyses of CO_2 variables have fairly good agreement with the true state overall (Figure 2.7(d) and Figure 2.8(d)). This result reveals that the multivariate data assimilation allows undesirable sampling error feedback between the CO_2 variables and the atmospheric variables. The atmospheric CO_2 is influenced only by the wind field and there is no relationship of CO_2 with temperature, humidity and surface pressure field in our nature run. Thus, the coupled error covariances between these irrelevant meteorological variables and CO_2 variables could make spurious values in the multivariate data assimilation, while the one-way multivariate data assimilation

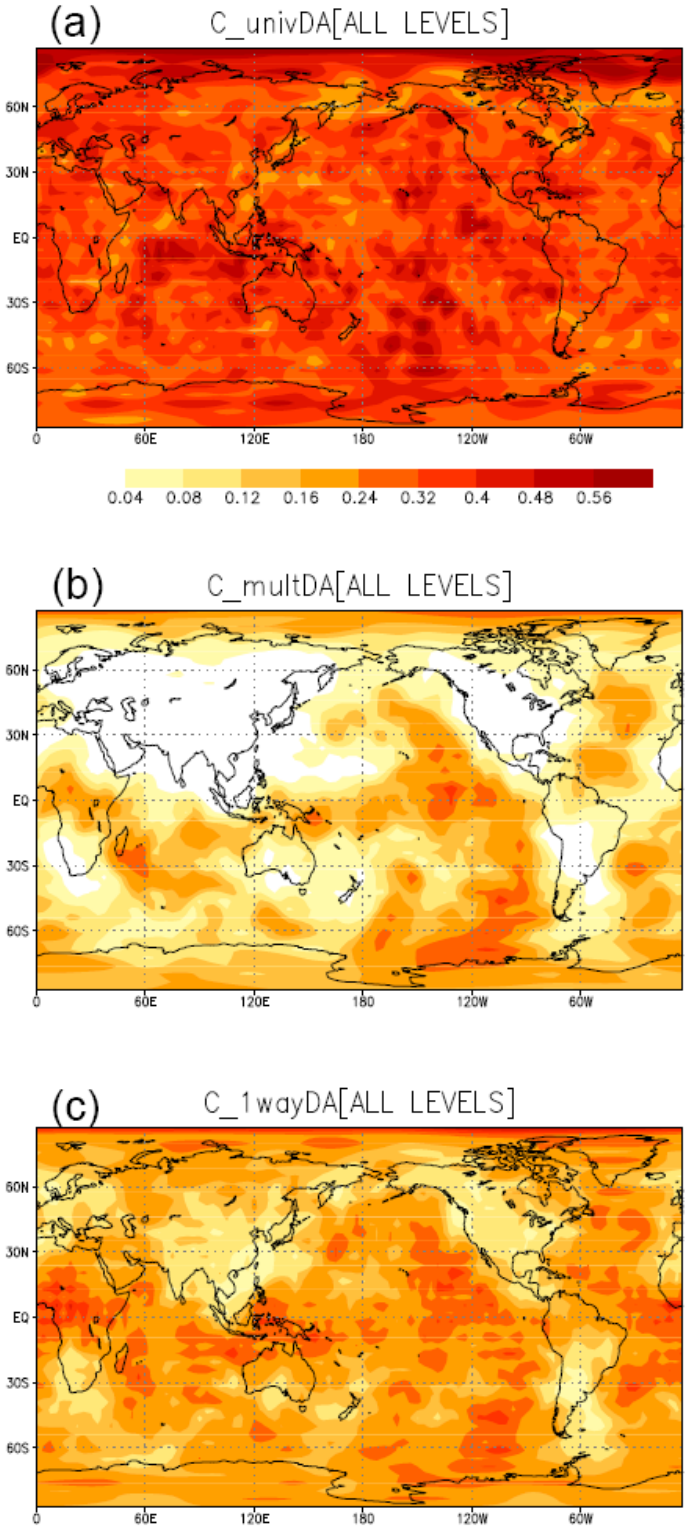

Figure 2. 10. Ensemble spread of atmospheric CO_2 analysis on the lowest layer of model from (a) the uncoupled data assimilation, (b) the multivariate data assimilation, and (c) the one-way multivariate data assimilation, after two months of analysis under ALL LEVELS experiments [unit: ppmv]

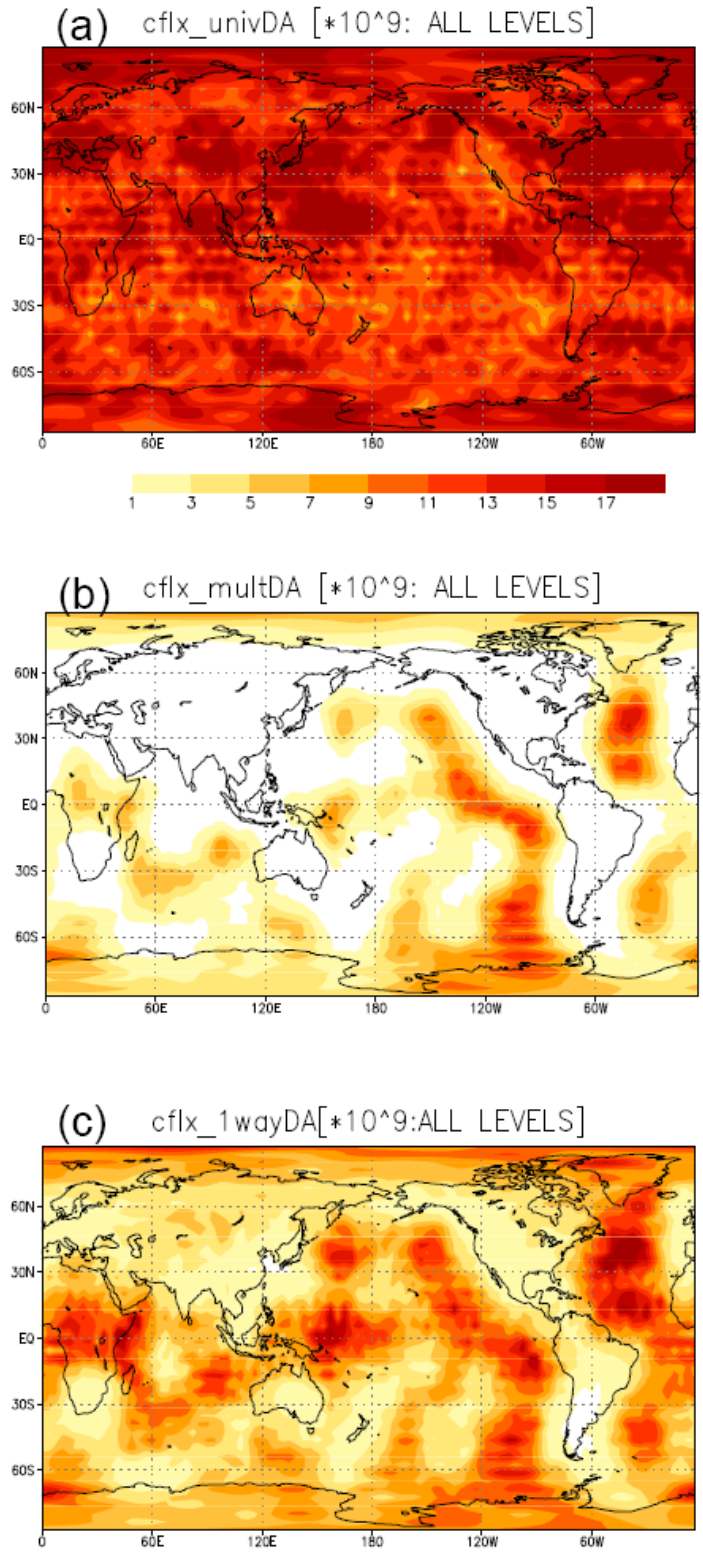

Figure 2. 11. Ensemble spread of CO_2 fluxes analysis at the surface from (a) the uncoupled data assimilation, (b) the multivariate data assimilation, and (c) the one-way multivariate data assimilation, two months of analysis under ALL LEVELS experiments [unit: 10^{-9}kg/m^2/s]

allows the wind field to give essential information to CO_2 variables and is not affected by their sampling errors.

Figure 2.10 and Figure 2.11 show the ensemble spread of the analysis for atmospheric CO_2 concentration and surface CO_2 fluxes. Since the uncoupled data assimilation diverges, the ensemble spread has a large value without overall physical meaning. On the other hand, both multivariate data assimilations have a larger spread over the ocean rather than over the land because the observations of atmospheric variables are based on the rawinsonde distribution, which is mainly over the land. In this context, the multivariate data assimilation has less spread over the land than the one-way multivariate because the uncertainties of CO_2 variables are linked to those of not only wind fields but also the other atmospheric variables in the multivariate data assimilation. In addition, the area where the atmospheric transport is active, over North Atlantic Ocean, is emphasized with the large spread of both atmospheric CO_2 and surface CO_2 fluxes.

2.5.2.2. “OCO + AIRS” experiment and “OCO” experiment

We examined only the one-way multivariate data assimilation for these two experiments since we found it was the optimal way to analyze CO_2 according to the results from ALL LEVELS experiments. Since we use the same observations of the atmospheric variables, and in the one-way multivariate approach the analysis of the standard atmospheric variables is not changed by assimilating atmospheric CO_2, the results of atmospheric variables in the analysis are the same in these two experiments as those shown in the Section 2.5.2.1. Figure 2.7 (e)-(f) and Figure 2.8 (e)-(f) show

that both of OCO+AIRS and OCO experiments have comparable results with the ALL LEVELS experiment in the one-way multivariate data assimilation scheme. The RMS error for the atmospheric CO_2 concentration becomes a little larger with 0.026 ppmv and 0.028 ppmv at the end of two-month analysis under OCO+AIRS experiment and OCO experiment, respectively, than that under the ALL LEVEL experiment. This is really a minor degradation when one considers the observation error of 1.0 ppmv. The accuracy of surface CO_2 fluxes is not degraded visibly either. Indeed, the RMS error is only 2.2×10^{-10} kg/m^2/s larger in OCO experiment than the ALL LEVEL experiment after two months of data assimilation. This means that the observation of CO_2 concentration near the surface plays a very important role in estimating surface CO_2 fluxes.

From the little difference between the results of "OCO+AIRS" and "OCO" experiments, one may conclude that AIRS CO_2 observations may not be useful for estimating surface CO_2 fluxes. However, this may be also due to the systematic shortcomings of the forecast model we used. Since SPEEDY-C has only seven vertical layers and the parameterization of convection scheme is relatively simple, the forecast model can underestimate the potential impact of AIRS CO_2 retrievals, which have a strong sensitivity in the upper troposphere, not near the surface. Indeed, the simulated AIRS CO_2 observations do not significantly improve the atmospheric CO_2 analysis at the levels where the observations are not made in our experiments. With the realistic system of LETKF/CAM3.5, however, an experiment with AIRS CO_2 retrievals shows some improvement of analysis and forecast in the atmospheric CO_2 fields not only at the level of highest averaging kernel but also extending to other

vertical layers (Liu et al., 2009). CAM3.5 is more sophisticated model with higher resolution compared to SPEEDY-C: $2.5^{\circ} \times 1.9^{\circ}$ horizontal resolution and 26 vertical levels up to 3.5 hPa. Improved deep convection schemes (e.g., Neale et al. 2008) could also improve the vertical mixing of atmospheric CO_2 in the forecast so that the AIRS observation may constrain the atmospheric CO_2 better. Pak and Prather (2001) suggested that satellite observations of CO_2 in the upper troposphere could provide a major constraint for the net carbon fluxes over the tropical land within their inversion method. In summary, although AIRS CO_2 information has little effect on constraining the surface fluxes in the SPEEDY model, it may be more effective in a more advanced system.

Thus, from the "OCO+AIRS" and "OCO" experiments, we cannot conclude that the AIRS CO_2 retrieval is not able to improve the CO_2 analysis near the surface, but we only stress that the instrument which has higher sensitivity of CO_2 near the surface can be more useful for analyzing the atmospheric CO_2 near the surface and surface CO_2 fluxes.

2.6. Summary and discussions

First, we developed a model, SPEEDY-C, to simulate atmospheric CO_2 by modifying an atmospheric GCM of intermediate complexity, the SPEEDY model. We confirmed that the performance of SPEEDY-C in transporting carbon is reasonable compared to the results of the NCAR CCM. The comparison supports the use of SPEEDY-C for the OSSEs in this study given that it is a very fast model, and that we can address many questions about data assimilation methods for atmospheric

carbon and surface fluxes that would be unfeasible using a high-resolution, computationally expensive GCM.

We then investigated three types of analyses by building different groups of state vectors in the Local Ensemble Transform Kalman Filter (LETKF) formulation and testing them through OSSEs: carbon-univariate, multivariate, and one-way multivariate data assimilation. Multivariate CO_2 data assimilation experiments were performed for the first time, and the results indicate that multivariate EnKF assimilation is much more effective in estimating both atmospheric CO_2 and surface carbon fluxes, even in the absence of observations or prior estimations of surface fluxes. Of the two multivariate schemes applied here, one-way multivariate data assimilation has better results than the fully multivariate analysis because it permits the error statistics of only the relevant variables to interact with in terms of CO_2 analyses. According to the "OCO+AIRS" and the "OCO" experiments, it can be concluded that the surface CO_2 fluxes can be estimated reasonably if the atmospheric CO_2 concentration can be observed near the surface.

Chapter 3: Coupling SPEEDY-C to VEGAS with SLand

3.1. Introduction

Human activities have increased the emission of CO_2 into the atmosphere since the industrial revolution. About half of released CO_2 is absorbed at the surface by land or ocean, and the rest of it remains in the atmosphere (Figure 1.2). The amount of CO_2 uptake at the surface, however, has significant temporal variability with respect to the climate whereas the anthropogenic emission does not fluctuate much with the seasons, but rather increases monotonically. For example, one can easily see that the global CO_2 growth rate during El Niño (La Niña) years becomes larger (smaller) in Figure 1.2. This relation between the variability of atmospheric CO_2 and ENSO has been confirmed by many previous studies (Bacastow, 1976; Keeling and Revelle, 1985; Braswell et al., 1997; Rayner et al., 1999; Jones et al., 2001; Zeng et al., 2005). Thus, CO_2 absorption capacity of land and ocean causes this difference between the emission and the mean and variability of atmospheric CO_2 concentration, shown as the green shading in Figure 1.2. And the amount of uptake by land and ocean surface exhibits both seasonal and interannual variability, which is obviously related to the climate.

Thus, the response of land and ocean uptakes to a given climate condition is very important to project future climate in terms of how much land and ocean would uptake atmospheric CO_2 released from human activities, and how long these reservoirs can contain it. From a number of studies based on the ground based

measurements, the regional distribution of surface CO_2 fluxes had been estimated on continental scales and there has been progress in understanding the global and regional carbon cycles. But, we still need more detailed understanding of where and how much the atmospheric CO_2 sinks and releases by land and ocean take place even under the current climate. It is not a trivial problem to understand because the climate and the CO_2 exchange process over land and ocean are linked to each other in rather complex ways of interactions and feedbacks. In other words, the problem is highly nonlinear.

In the previous studies on the oceanic CO_2 uptake, many of them agree on relatively modest contribution of ocean to the variability of atmospheric CO_2 (Knorr, 2000; Bousquet et al., 2000; Feely et al., 2002; Rodenbeck et al., 2003). On the other hand, the atmospheric CO_2 uptake by the land surface is responsible for most of the variability of atmospheric CO_2 (Bousquet et al., 2000; Gurney et al., 2002; DeFries et al., 2002; Rodenbeck et al., 2003; Friedlingstein et al., 2006). The surface CO_2 fluxes over the land change with larger amplitudes depending on the climate conditions, whereas the oceanic CO_2 fluxes are considered to have less variability in the interannual time scales.

Based on this finding we decided to develop a coupled atmosphere-vegetation model in order to create a more realistic "nature" with time-varying surface CO_2 fluxes over land. Therefore, we coupled the SPEEDY-C model with the dynamic terrestrial carbon model VEgetation-Global-Atmosphere-Soil, VEGAS, (Zeng et al, 2005), which is turn coupled to the physical land surface model Simple-Land, SLand (Zeng et al., 2000). With this system, we expect to simulate surface CO_2 fluxes that

evolve seasonally and interact with climate anomalies. For the oceanic CO_2 fluxes, we have used the prescribed monthly means estimated by Takahashi et al. (2002) with a global mean rate of -2 PgC/yr. It is reasonable assumption because the variability of surface CO2 flux is dominant over the land on the interannual time scale.

This coupled system should not only create a more realistic nature run with variable surface carbon fluxes, but also make running an imperfect model OSSE possible. In a simulation mode of assimilation experiments, we can use the coupled system as the nature while the uncoupled SPEEDY-C model continues to be used for the ensemble forecast. Thus, this coupled system allows us to deal with the model error and to see the performance of data assimilation in the case where the model error is one of the serious problems, as it is in the real world.

Section 3.2 describes the way we coupled SPEEDY-C with VEGAS and SLand, and there is a short summary on a spin-up run in Section 3.3. Results are shown in Section 3.4 and we summarize and discuss the coupling work in Section 3.5.

3.2. Methods

3.2.1. Interface among atmosphere, vegetation, and land models

SPEEDY-C is coupled with VEGAS-SLand as described in schematic Figure 3.1. Details of each model are described in Zeng et al. (2000) for SLand, Qian (2008) for VEGAS, and Molteni (2003) for SPEEDY. In this section, the interface among

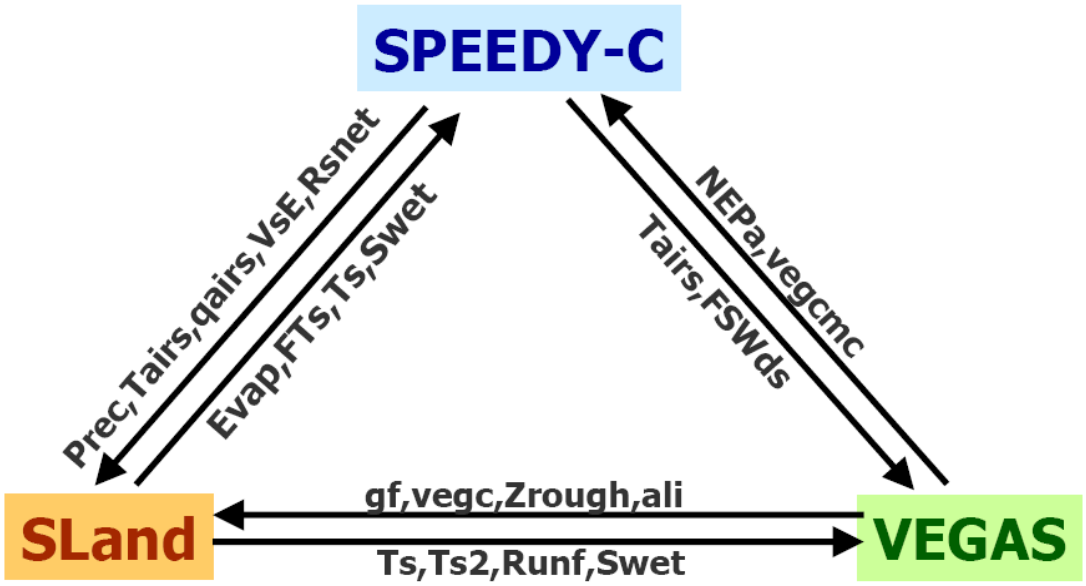

Figure 3. 1. Schematic diagram of interface among the coupled components of atmosphere (SPEEDY-C), vegetation (VEGAS), and land surface (SLand). Variables in the interface are described in section 3.2.1. (Prec-precipitation, Tairs-temperature near the surface, qairs-specific humidity near the surface, VsE-wind speed near the surface, Rsnet-net shortwave radiation at the surface, Evap-evaporation, FTs-sensible heat, Ts-surface temperature at layer 1 (top layer), Swet-soil wetness, Ts2-soil temperature at layer 2, Runf-runoff, gf-growth factor, vegc-vegetation cover, Zrough-roughness length, ali-leaf area index, FSWds-downward shortwave radiation at the surface, NEPa-surface CO_2 fluxes between atmosphere and land, vegcmc-annual mean of vegetation cover)

the coupled components necessary to develop a coupled system is discussed. First, SPEEDY-C gives SLand the information of precipitation, the temperature, specific humidity, and wind speed near the surface, and the net shortwave radiation at the surface. In turn, SLand provides SPEEDY-C with evaporation, sensible heat, surface temperature, and soil wetness over the land. We still calculate the surface fluxes over ocean with the formulation in the original version of SPEEDY because VEGAS-SLand is designed only for the land surface processes. In the interface between SLand and VEGAS, SLand provides VEGAS with soil temperature, runoff, and soil wetness while VEGAS returns the CO_2 dependent growth factor, vegetation cover, roughness

length, and leaf area index to SLand. Lastly, in the interface between atmosphere and vegetation, SPEEDY-C provides temperature near surface and downward shortwave radiation at the surface to VEGAS, and VEGAS calculates the surface CO_2 fluxes between the atmosphere and vegetation and updates the vegetation cover annually. In order to accelerate the spin-up process, the time step of the vegetation model is set up as one day while SPEEDY and SLand have the same time step of 20 minutes. Thus, the input variables of atmosphere and land models to the VEGAS are averaged at every 00Z daily.

3.2.2. Additional boundary conditions

Since VEGAS-SLand requires the boundary forcing of topographic gradient and ice cover, those fields are obtained by interpolating a fine-grid data (GLOBE task team, 1999; Peltier, 1994) to the SPEEDY-C resolution (Figure 3.2).

3.2.3. Adjustment of land-sea mask

SPEEDY has a fractional land-sea mask whereas VEGAS-SLand use an integer mask. Both models use unity for the land and zero for the ocean surface. But SPEEDY has fractional numbers around the coastline between land and ocean, proportional to the ratio of land and ocean in terms of area. Thus, we made VEGAS-SLand accept any point where the land-sea mask of SPEEDY is not zero as a land surface point.

Furthermore, SLand computes the variables only over land and VEGAS only over land without ice so that the index of land for SLand computation and that for

VEGAS have been calculated according to the land-sea mask computed in a way described above and ice-cover boundary information. Then, these indices are given for the input of VEGAS and SLand systems to recognize whether the grid point belongs to land with no ice, ocean, or ice-covered land.

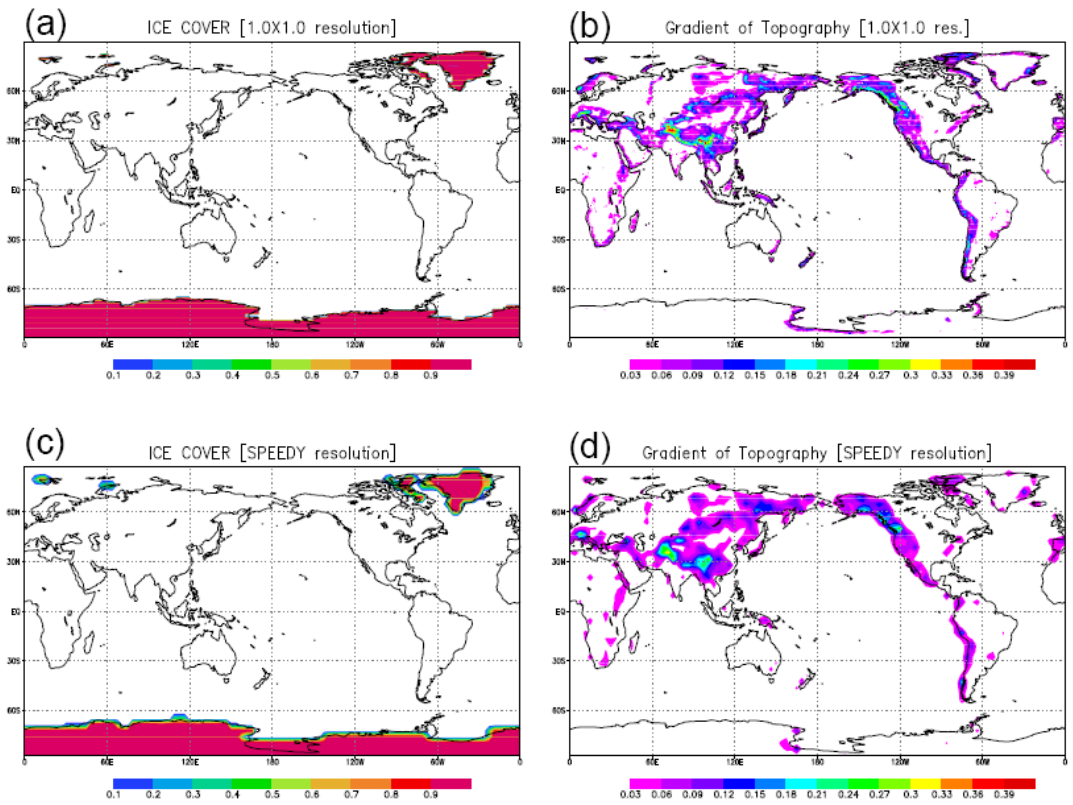

Figure 3. 2. Additional boundary forcing for VEGAS: (a) ice cover data with 1°×1° resolution (Peltier, 1994), (b) a gradient of topography data with 1°×1° resolution (GLOBE task team, 1999), (c) ice cover data interpolated to SPEEDY-grid T30 resolution, and (d) a gradient of topography interpolated to SPEEDY grid system.

3.2.4. Soil moisture and tropical rainfall over land

The original version of SPEEDY uses a seasonally changing climatology for the soil moisture fields. Before performing the coupling, we compared the prescribed soil wetness of original SPEEDY with that from SLand-VEGAS offline run forced by

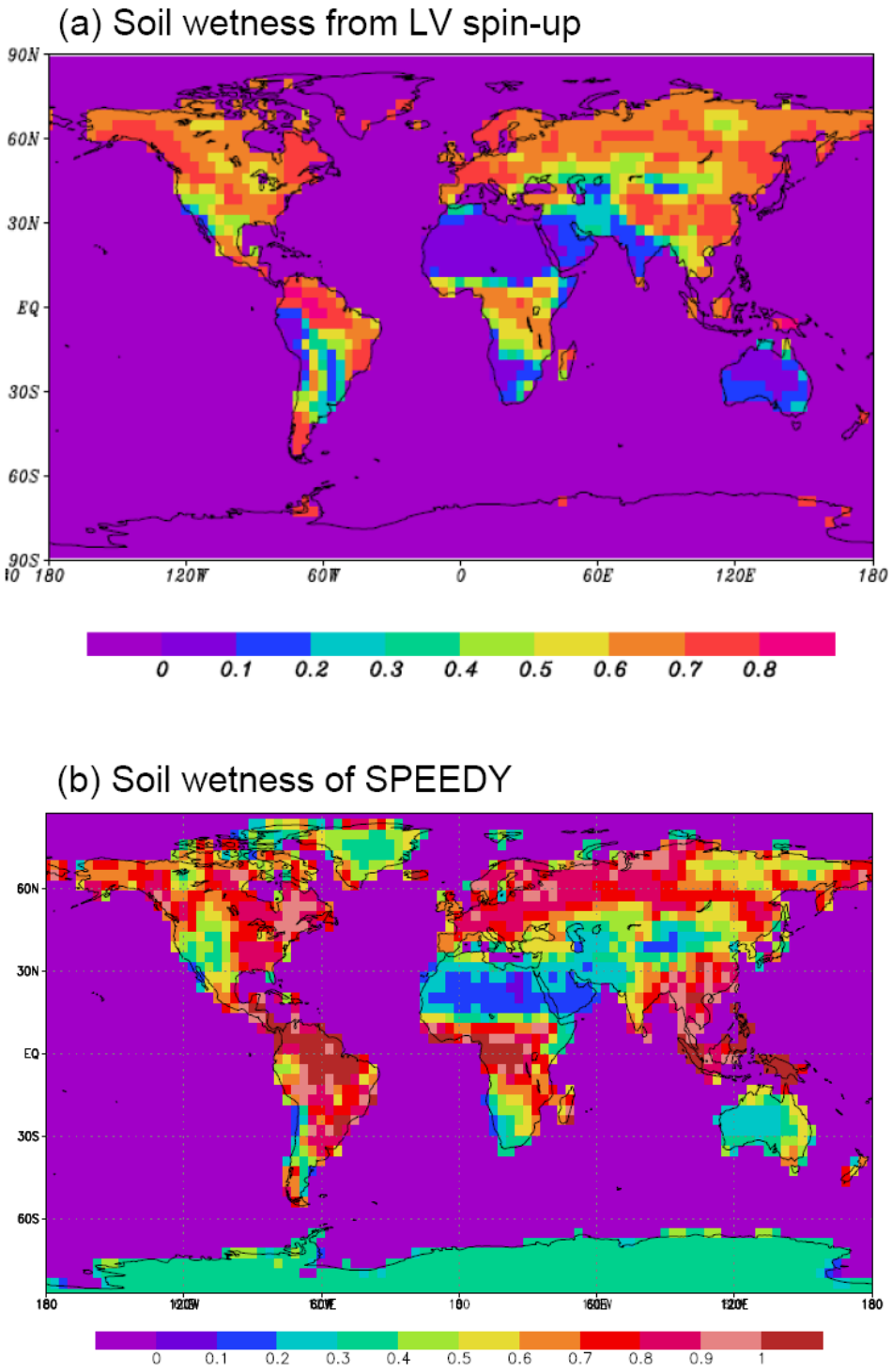

Figure 3. 3. Annual mean of soil wetness in (a) VEGAS-SLand (LV) offline simulation forced by SPEEDY climatology, (b) the prescribed boundary condition used in the original version of SPEEDY. (unit: dimentionless)

a climatology of atmospheric variables from SPEEDY, and found that the soil wetness of SPEEDY is significantly different from that of SLand (Figure 3.3). By definition, soil wetness (Swet) is the relative soil water saturation as a ratio of modeled soil moisture (mm) to the maximum value, which is 500 mm in SLand and 350mm in SPEEDY, so that Swet varies from 0 to 1. The prescribed boundary forcing in SPEEDY has highly saturated soil moisture over the tropical land whereas SLand-Vegas has a maximum value of soil wetness less than unity. Zeng et al. (2008) validated the modeled soil moisture by SLand on seasonal, interannual and longer timescales. However, it turned out that the SPEEDY is tuned in such a way that it requires large areas where the soil moisture is saturated in order to maintain a realistic precipitation over the tropical land, but this large soil moisture is not realistic. Indeed, with the modeled soil moisture calculated by SLand, the fully coupled atmosphere-vegetation-land model did not have rainfall over the tropical land and those areas became dry like deserts. Thus, we had to tune the modeled soil moisture from SLand before giving it to the atmosphere and vegetation components. We made the soil wetness saturated over the tropics to let SPEEDY produce realistic precipitation by multiplying the soil moisture with Gaussian weights dependent of latitude between -20°S and 20°N. This resulted in a pattern of soil moisture similar to the original SPEEDY climatology and produced reasonable ranges of rainfall over the tropical land.

3.3. Spin-up

It is necessary to obtain an equilibrium state of the coupled system through the spin-up run. Here, it is assumed the equilibrium state is reached when the annual mean of NEP (Net Ecosystem Productivity) converges to zero. Because the SLand-VEGAS is computationally economic compared to SPEEDY and we need to run the soil and vegetation models for hundreds of years in order to have the convergence of vegetation and soil variables to the proper states, an offline spin-up run of SLand-VEGAS was done first with the atmospheric forcing of SPEEDY climatology and then we run a fully coupled system.

We ran the SPEEDY for nine years to get seasonally varying climatology of variables such as precipitation, temperature, specific humidity, wind, and radiation at the surface, which were then used to force the Vegetation-Land model. Under the given SPEEDY climatology, the SLand-VEGAS offline simulation continued for 600 years. During the first 200 years, an "accelerator" factor was used to help the soil carbon pool reach the equilibrium state relatively fast. From the late states of the variables of SLand-VEGAS offline run, we let the fully coupled system run for 30 years where the prescribed ocean fluxes of CO_2 have been included with a magnitude of -2 PgC/yr (Takahashi et al. 2002) and with no anthropogenic fluxes.

3.4. Results

3.4.1. Offline Land-Vegetation spin-up run

Figure 3.4 shows that the vegetation-land model has reached an equilibrium state of Gross Primary Productivity (GPP), Net Primary Productivity (NPP), Net Ecosystem Productivity (NEP), Carbon in vegetation, and carbon in soil. The global total GPP converges to 110 PgC/yr, NPP to 55 PgC/yr, vegetation carbon pool to 550 PgC, and soil carbon to about 2000 PgC. These are comparable with those resulted from Zeng et al. (2005) which includes the experiment forced by the observed record of real atmospheric variables: The global total GPP is 122 PgC/yr, NPP is 58 PgC/yr, vegetation carbon pool is 550 PgC, and soil carbon is 1850 PgC. Figures 3.5 and 3.6 show the spatial pattern of NEP, NPP, GPP, respiration, carbon in different vegetation pools, and soil resulting from the offline spin-up run. These results are also reasonable compared to the results in Zeng et al. (2005) which used the same vegetation and land model with real atmospheric observations.

3.4.2. Fully coupled atmosphere-vegetation-land spin-up run

The global total GPP is about 90 PgC/yr, NPP about 40 PgC/yr at the end of 30-year run. NEP converges to zero annual mean (Figure 3.7). Compared to the result of NPP from the offline simulation, NPP from the coupled spin-up simulation has a relatively smaller value. We could find the reason of this from the difference of precipitation over land between the offline simulation and coupled run (Figure 3.8). The climatology of SPEEDY used for the forcing in the offline spin-up has generally larger precipitation over land than the precipitation calculated by interactive mode of

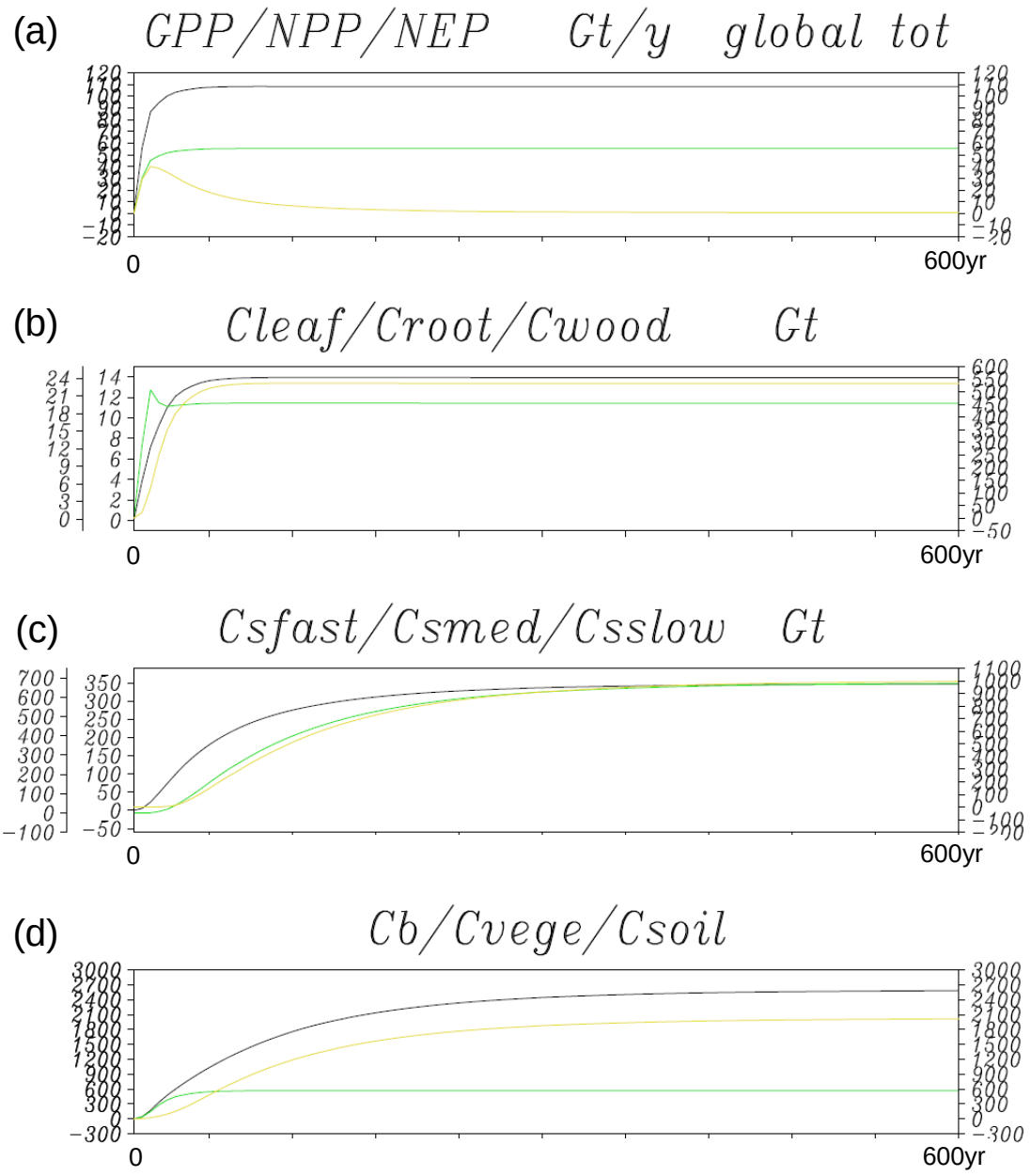

Figure 3. 4. Time series of global total of major variables in VEGAS during offline spinup simulation with SPEEDY climatology: (a) GPP (Gross Primary Productivity: black), NPP (Net Primary Productivity: green), and NEP (Net Ecosystem Productivity: yellow), (b) Cleaf (leaf carbon: black), Croot (root carbon: green), and Cwood (wood carbon: yellow), (c) Csfast (fast soil carbon:black), Csmed (intermediate soil carbon:green), and Csslow (slow soil carbon: yellow), (d) Cb (total biosphere carbon, i.e. soil carbon + vegetation carbon: black) , Cvege (vegetation carbon: green), and Csoil (soil carbon: yellow). Values are averaged annually.

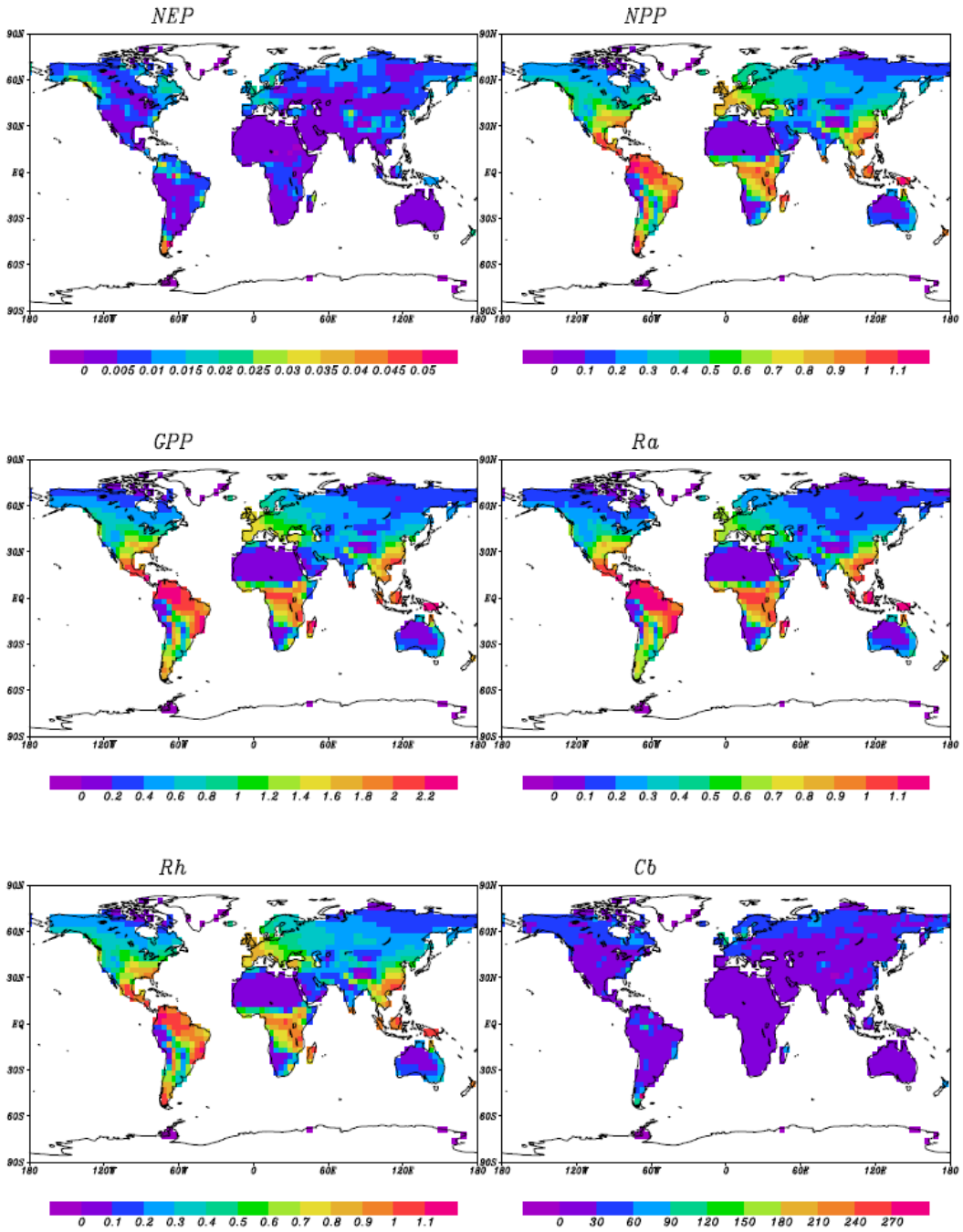

Figure 3. 5. Annual mean fields of NEP (kg/m^2/yr), NPP (kg/m^2/yr), GPP (kg/m^2/yr), Ra (autotrophic respiration, kg/m^2/yr), Rh (heterotrophic respiration, kg/m^2/yr), and Cb (kg/m^2) for the last year of 600-year Land-Vegetation offline spin-up.

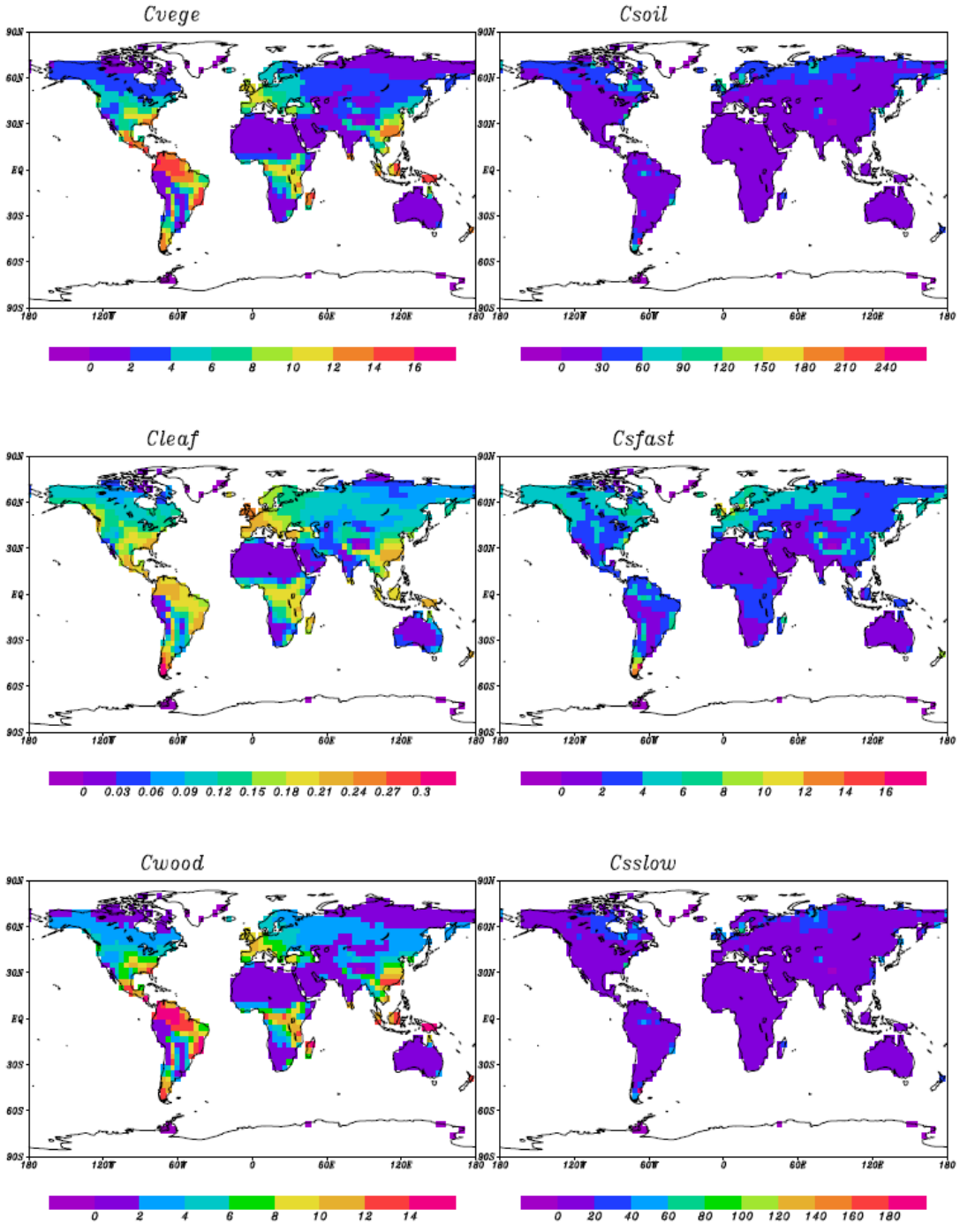

Figure 3. 6. Same as Figure 3.5 except for Cvege, Csoil, Cleaf, Csfast, Cwood, and Csslow (unit: kg/m^2).

coupled simulation. Especially over the southern America, there is large dry area in the coupled simulation compared to the climatology of SPEEDY. Thus, this environment in the coupled system could not have as much vegetation as the atmospheric condition with SPEEDY climatology so that NPP converged to a lower level in the coupled spin-up simulation than in the offline run.

Vegetation carbon pool has a value of 380 PgC, and soil carbon is about 1780 PgC after 30 years. Figure 3.9 shows the spatial distribution of NEP in each season. In general, the results indicate that the vegetation uptakes atmospheric CO_2 during growing season whereas the land surface releases CO_2 into the atmosphere during the vegetation decaying season. Although it is necessary to deal with details of the results more carefully in order to use the nature run for other applications, for our purposes we intend to use the coupled system in a simulation where the only requirement is that the "nature" produce surface carbon fluxes with a reasonable seasonal variability, and that the "forecast" model be significantly different, as it happens in the real world, and the coupled system that we developed satisfies this requirement.

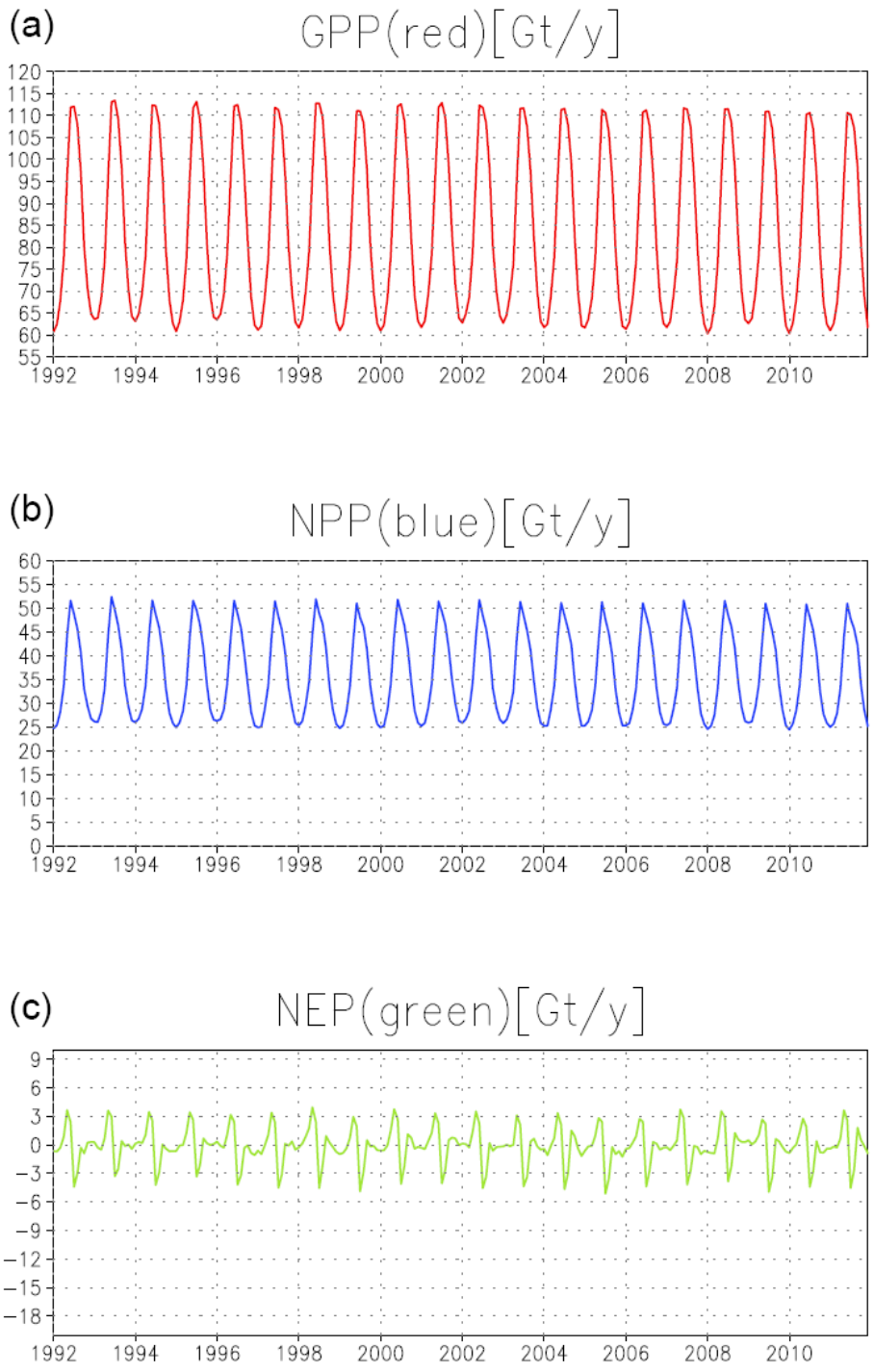

Figure 3. 7. Time series of monthly mean of global total (a) GPP, (b) NPP, and (c) NEP for last 20 years' fully coupled spin-up.

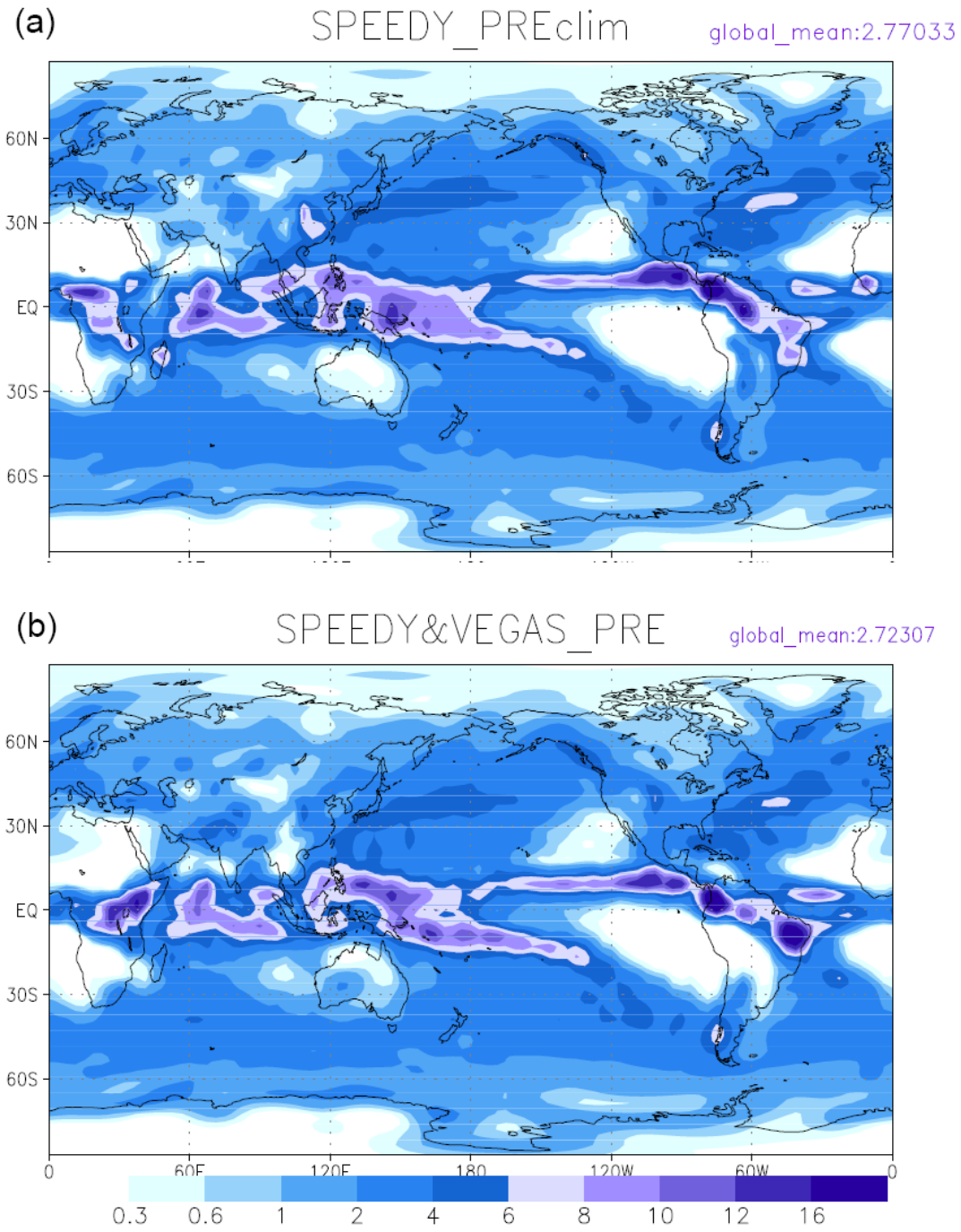

Figure 3. 8. (a) Annual mean of precipitation from SPEEDY climatology (nine-year mean) and (b) last 10-year mean precipitation of coupled SPEEDY-VEGAS-SLand simulation for 30 years (unit: mm/d).

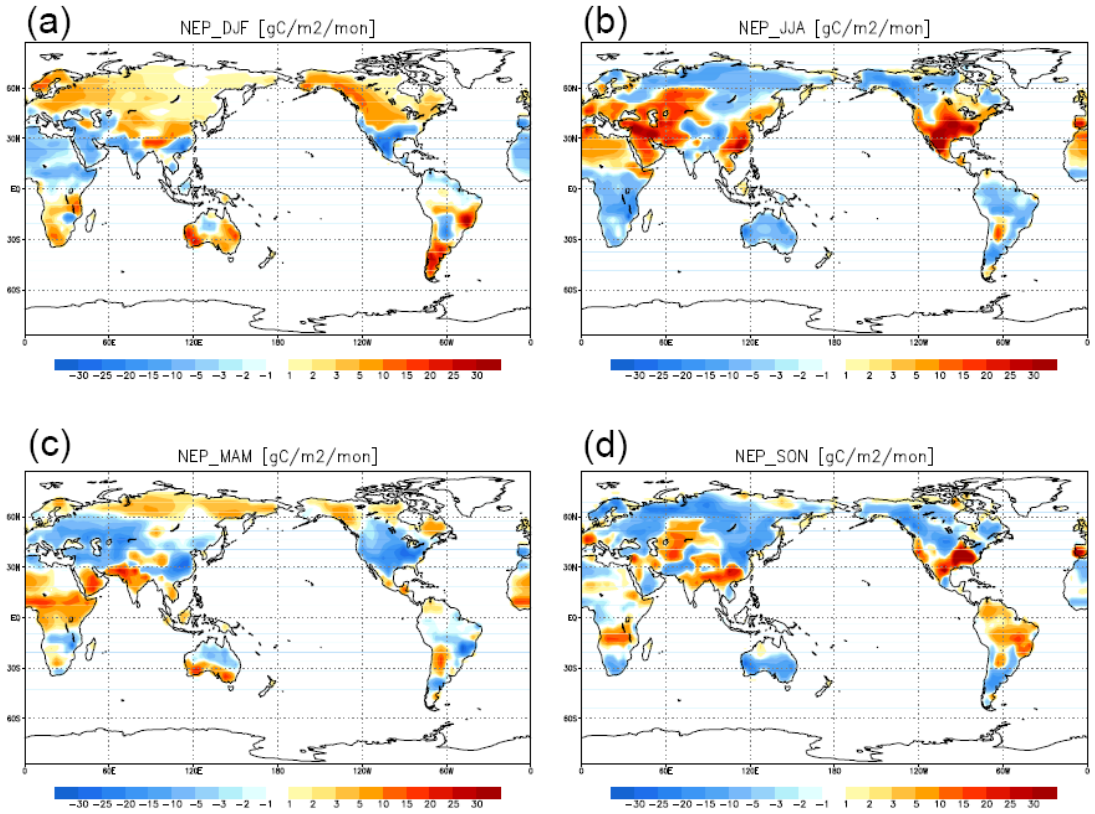

Figure 3. 9. Seasonal mean of NEP (surface carbon flux) in (a) DJF, (b) JJA, (c) MAM, and (d) SON of last year in the fully coupled atmosphere-vegetation-land spin-up. (positive: carbon sources, negative: carbon sinks)

3.5. Summary

We coupled SPEEDY-C with a terrestrial dynamic carbon model (VEGAS) and a simple physical land model (SLand). Among the components of atmosphere, land, and vegetation, the heat, water, and energy fluxes are obtained through coupling. Then, we made a spin-up run to get an equilibrium state of land and vegetation with the SPEEDY-C atmosphere through an offline simulation of land and vegetation and a fully coupled simulation of three components in order. Since the coupled model produces reasonable CO_2 fluxes over land, it is possible to use it as

"nature", since it has a more realistic carbon cycle with time-varying CO_2 fluxes which we will try to estimate through EnKF data assimilation using the SPEEDY-C model. The coupled system can also be used for other climate studies associated with CO_2 and dynamic vegetation, although users must be aware that the soil moisture was tuned in the tropics.

Chapter 4: Imperfect Model Simulation: Bias Correction, Adaptive Inflation and Estimation of Observation Errors

4.1. Introduction

In Chapter 2, we have seen the performance of LETKF on estimating surface CO_2 fluxes under the perfect model scenario. Without any direct observation or *a-priori* information of surface carbon fluxes, data assimilation could produce a reasonable estimate of these fluxes in the multivariate analysis system. Now, it is necessary to consider a more realistic case in which the model error cannot be ignored. Chapters 2 and 3 show we can use for this purpose two different models: the SPEEDY-C and the SPEEDY-VEGAS. Thus, we are able to do OSSEs (Observing System Simulation Experiments) under the imperfect model assumption by using SPEEDY-C for the forecast model and SPEEDY-VEGAS for the nature run.

In order to deal with model errors, we applied two additional techniques: a bias correction and an adaptive inflation. First, we implemented a simple model bias correction which is similar to that introduced by Danforth et al. (2007). Next, we applied an advanced method, a simultaneous estimation of adaptive inflation and observation errors, introduced by Li et al. (2009, hereafter referred as LI09). With this method, we could estimate an adaptive inflation as well as the observation errors within the data assimilation ("online"). Both methods are introduced in Section 4.3

which also includes discussions on the difficulties of estimating inflation for a variable for which there are no remote observations such as surface CO_2 fluxes.

This chapter is organized as follows: The experimental design is discussed in Section 4.2. Section 4.3 describes a method to correct the model bias and shows the results from that. Next, an adaptive inflation technique applied to our case is presented in Section 4.4 and the results are shown. Lastly, Section 4.5 has a summary of this chapter.

4.2. Experimental Design

The coupled atmosphere-land-vegetation model, SPEEDY-VEGAS, introduced in Chapter 3 is now used for the nature run while the SPEEDY-C in Chapter 2 is used for the ensemble forecast. Since the analysis occurs every six hours, we forced the nature run to update surface CO_2 fluxes over the land every six hours by changing the time step of vegetation model from one day to six hour when we created the nature run. In the nature run, CO_2 fluxes over land are calculated by the coupled model, whereas those over the ocean are prescribed monthly (Takahashi et al., 2002) with a rate of -2 PgC/yr. In the imperfect model simulation, there is no fossil fuel emission in the nature run.

Initial conditions of (U, V, T, q, Ps, CO_2) for the 20-ensemble forecast were created by adding 20 random perturbations to the 20 states which were chosen randomly in time from the one-year nature run, as in the perfect model simulation. For the surface CO_2 fluxes, the states at another 20 timesteps from the nature run were chosen, and random perturbations were added to them. The magnitudes of the

random perturbations added to (U, V, T, q, Ps, CO_2) are the same as those used for the perfect model experiment in Chapter 2. For the surface CO_2 fluxes, a standard deviation of the random perturbation is 1.0×10^{-10} kg/m2/s. Again, the initial conditions of surface CO_2 fluxes do not have any *a-priori* information as shown in Figure 4.1.

For the imperfect model simulation, we set a fixed multiplicative inflation of 8% for all the dynamic variables for the control run and the bias correction experiment without an adaptive inflation technique. Later, for the experiments including adaptive inflation and observation errors estimation, the initial guesses of observation errors were given as twice the true values (Table 4.1), and the inflation started from 10% at the initial time. The way observations of (U, V, T, q, Ps, CO_2) were generated is same as the case of perfect model simulation (Chapter 2). Here we only examined "ALL LEVEL" experiment described in Section 2.4.1, and only the 1-way multivariate data assimilation has been examined since we found it had an optimal performance in Chapter 2.

4.3. Bias correction

4.3.1. Methodology

Under the imperfect model simulation, one can expect a significant difference of climatologies between the forecast run and the nature run. With an imperfect model, as when the true atmosphere is compared with a model, the difference in climatology is not associated with forecast errors due to the errors in the initial

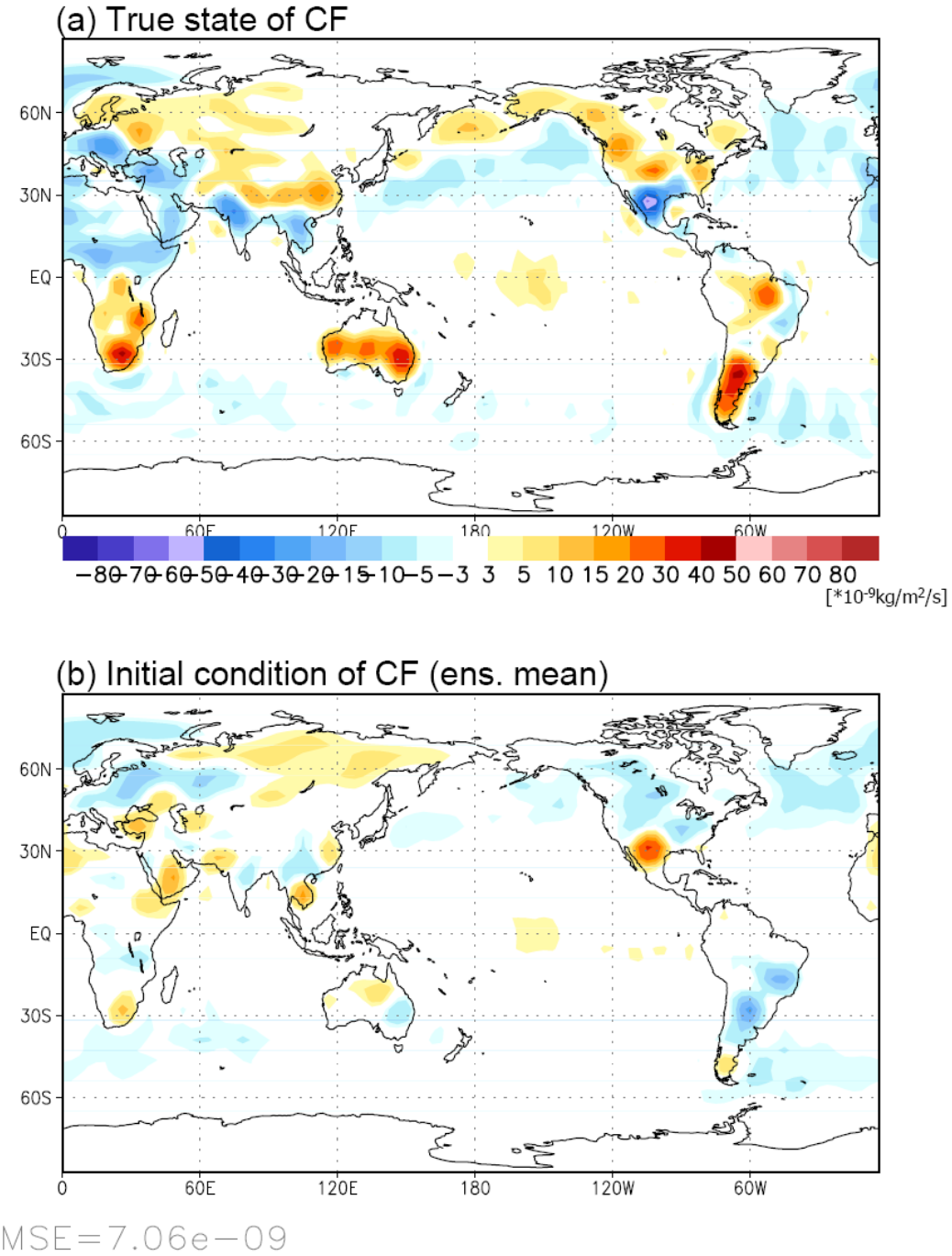

Figure 4. 1. (a) True state of surface CO_2 fluxes at the initial time step, (b) initial condition for surface CO_2 fluxes. (unit: $10^{-9}kg/m^2/s$). (positive: CO_2 sources, negative: CO_2 sinks)

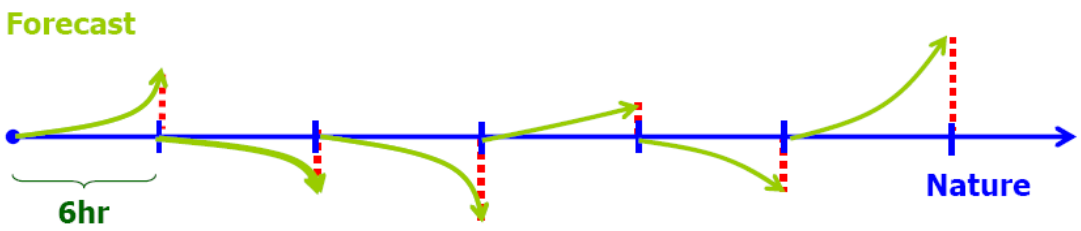

Figure 4. 2. Schematic plot to describe the low-dimensional correction of model bias: blue arrow stands for nature (or reanalysis), and green arrows indicate every 6 hour forecast stating from the nature run. The departures of 6-hour forecasts from the nature run happen to be caused by the discrepancy between the forecast and the nature runs. The two-month averaged field of those departures is considered as the model bias, and hence it is subtracted from the background at every analysis step.

condition, but it indicates a presence of systematic errors in the forecast model. It is caused by the model deficiencies such as inaccurate forcings and parameterizations (Danforth et al., 2007). We applied the method of low-dimensional correction in Danforth et al. (2007) for estimating and correcting model bias. Figure 4.2 schematically describes the simple way model bias is estimated. We made a series of 6 hour forecasts which restart from the initial conditions of the nature run every six hours. Thus, the time average of these departures can be considered as an estimate of the forecast bias. We calculate this model bias over a period of two months. The model bias is then subtracted from the 6 hour forecast before every analysis step. The bias correction is applied to only the atmospheric variables, not to the CO_2 variables. We do not know in the real world the truth ("nature"), but the extensive experience with atmospheric data assimilation and the reanalyses data sets assure we have good enough estimates of the 4D-state of the atmospheric variables u, v, T, q and ps to correct their biases with this method. On the other hand we do not yet have "CO_2 reanalyses" that we could use to correct the model bias of CO_2 forecasts.

4.3.2. Results from the bias correction experiment

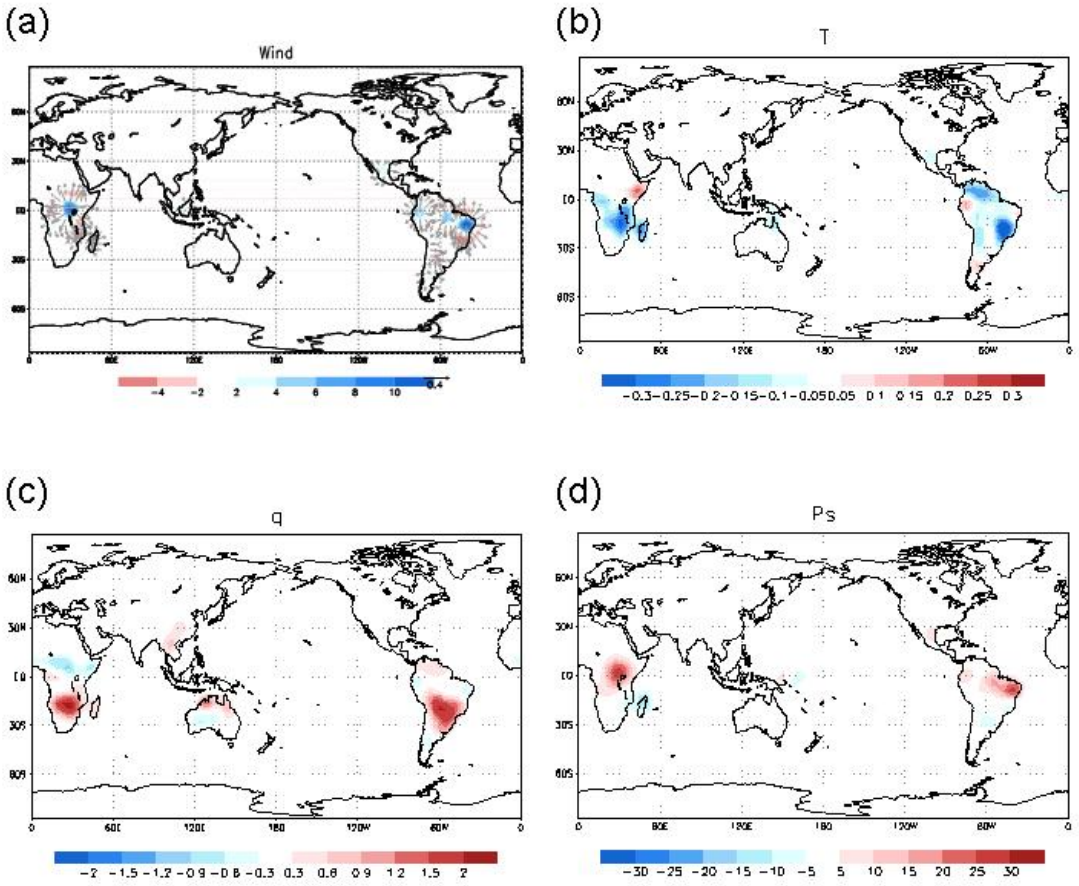

Figure 4. 3. Estimated model bias from the low-dimensional correction in (a) wind (m/s) (shading: divergence, unit: 10^{-7}/s), (b) T(K), (c) q (g/kg) on the lowest layer, and (d) surface pressure (Pa), for two months of analysis period (model minus nature, positive: forecast overestimates, negative: forecast underestimates)

Figure 4.3 shows that major features of the model bias appear over the tropics where soil moisture fields were increased in the SPEEDY-VEGAS over the tropical land in the coupled system in order to improve the precipitation pattern as described in Section 3.2.4. The differences in soil moisture dominate the differences between the nature run. The forecast model has biases of relative divergence and higher surface pressure over the equatorial land compared to the nature run. Moreover, there

is lower temperature and higher precipitation over the land near 20°S. If one compares the climatologies of soil moisture, evaporation, and precipitation in the forecast model with those in the nature run (Figure 4.4), the model bias can be explained as follows: High (low) soil moisture on the equatorial land ignites the strong (weak) evaporation and convection, which corresponds to the strong (weak) convergence on the surface layer. That causes more (less) precipitation over the area. The strong (weak) evaporation can explain the low (high) temperature, high (low) humidity on the lowest layer.

After correcting for this estimated model bias, we obtained a remarkable improvement in the analysis of variables that have observations such as wind, temperature, humidity and even atmospheric CO_2 concentration (Figure 4.5). It is encouraging that we obtained an improvement on the atmospheric CO_2 analysis without correcting its bias. This is because the atmospheric CO_2 transport is predicted better after the model bias of wind is corrected. Since the atmospheric CO_2 is also linked to the surface CO_2 fluxes, however, the analysis of the atmospheric CO_2 can get a negative effect from a poor analysis of the surface CO_2 flux fields. Indeed, the analysis of surface CO_2 fluxes still diverged with time even though they are better due to the indirect effect from the bias correction of wind. Figure 4.6 shows the spatial distribution of analysis errors in the zonal wind, the atmospheric CO_2 on the lowest layer, and the surface CO_2 flux fields without and with bias correction. Bias correction clearly gets rid of the most of the mean analysis errors in the wind fields. Other meteorological variables also benefit from the bias correction remarkably as

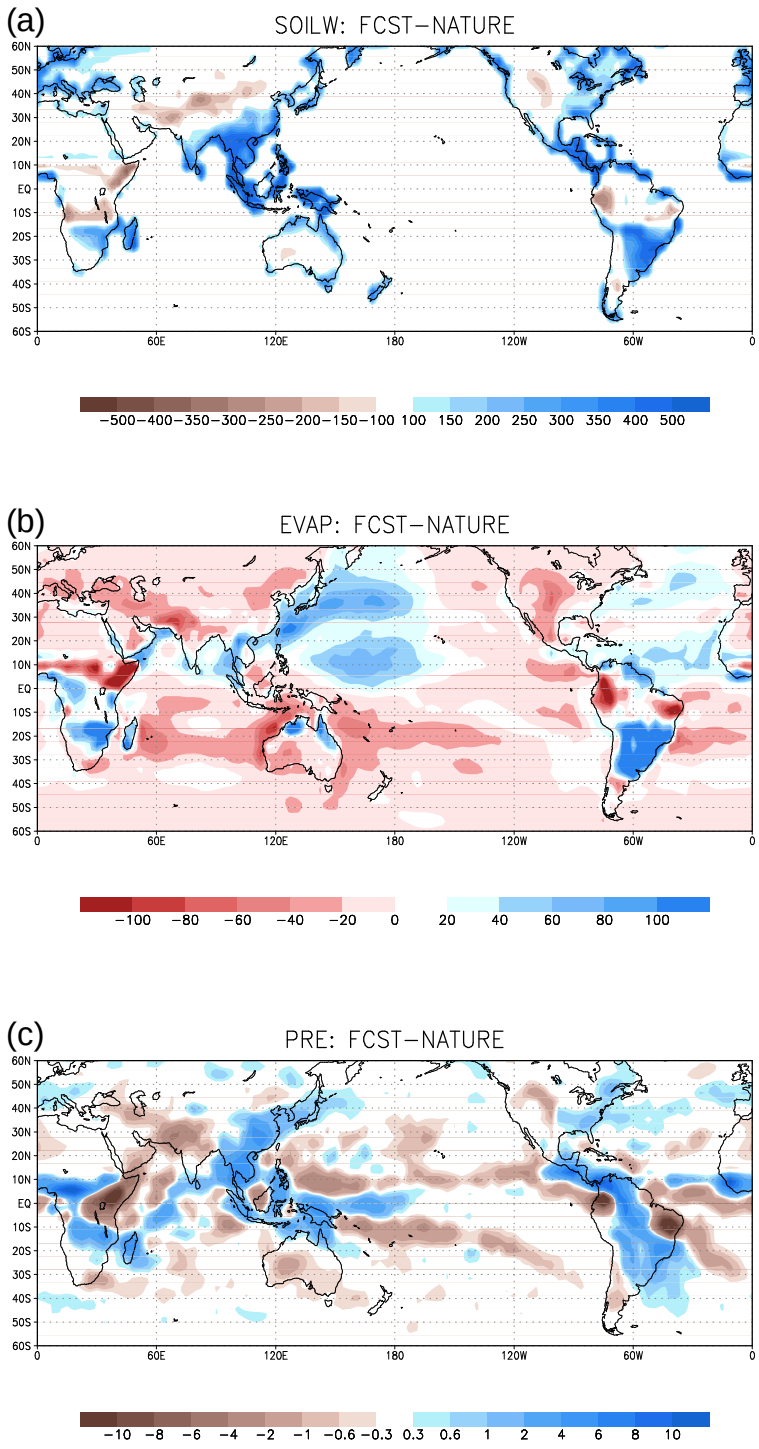

Figure 4. 4. Difference of the climatologies between the forecast model (SPEEDY-C) and the nature run (SPEEDY-C coupled with VEGAS-SLand); (a) soil moisture (mm), (b) evaporation (W/m^2), and (c) precipitation (mm/d). (positive: the forecast has larger values than the nature run, negative: the forecast has less values than the nature)

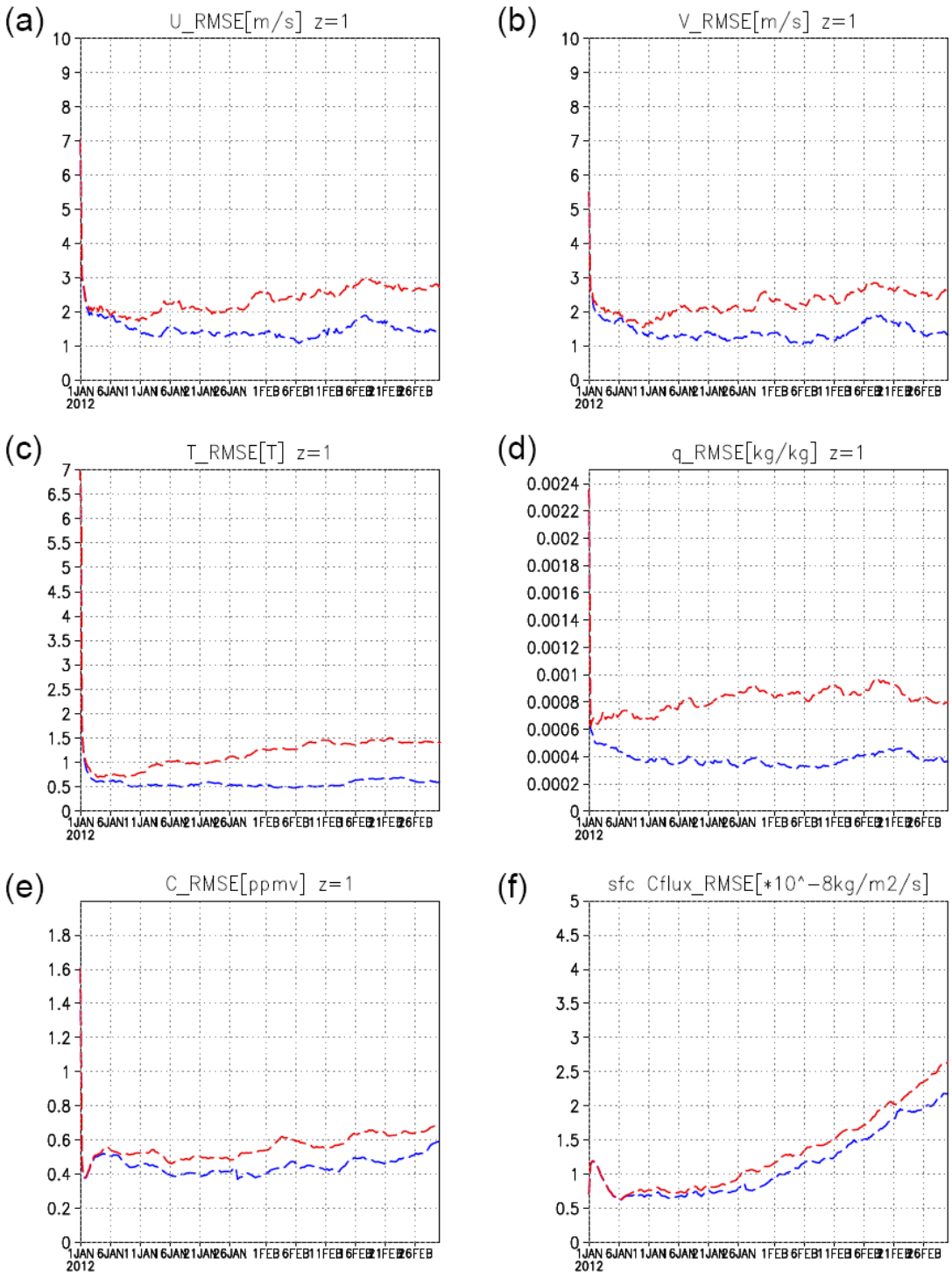

Figure 4. 5. RMS errors of (a) U (m/s), (b) V (m/s), (c) T (K), (d) q (kg/kg), (e) atmospheric CO_2 (ppmv) on the bottom layer, (f) surface CO_2 fluxes (10^{-8} kg/m^2/s) in the analysis. Red lines indicates the control run, and blue lines result from the bias correction.

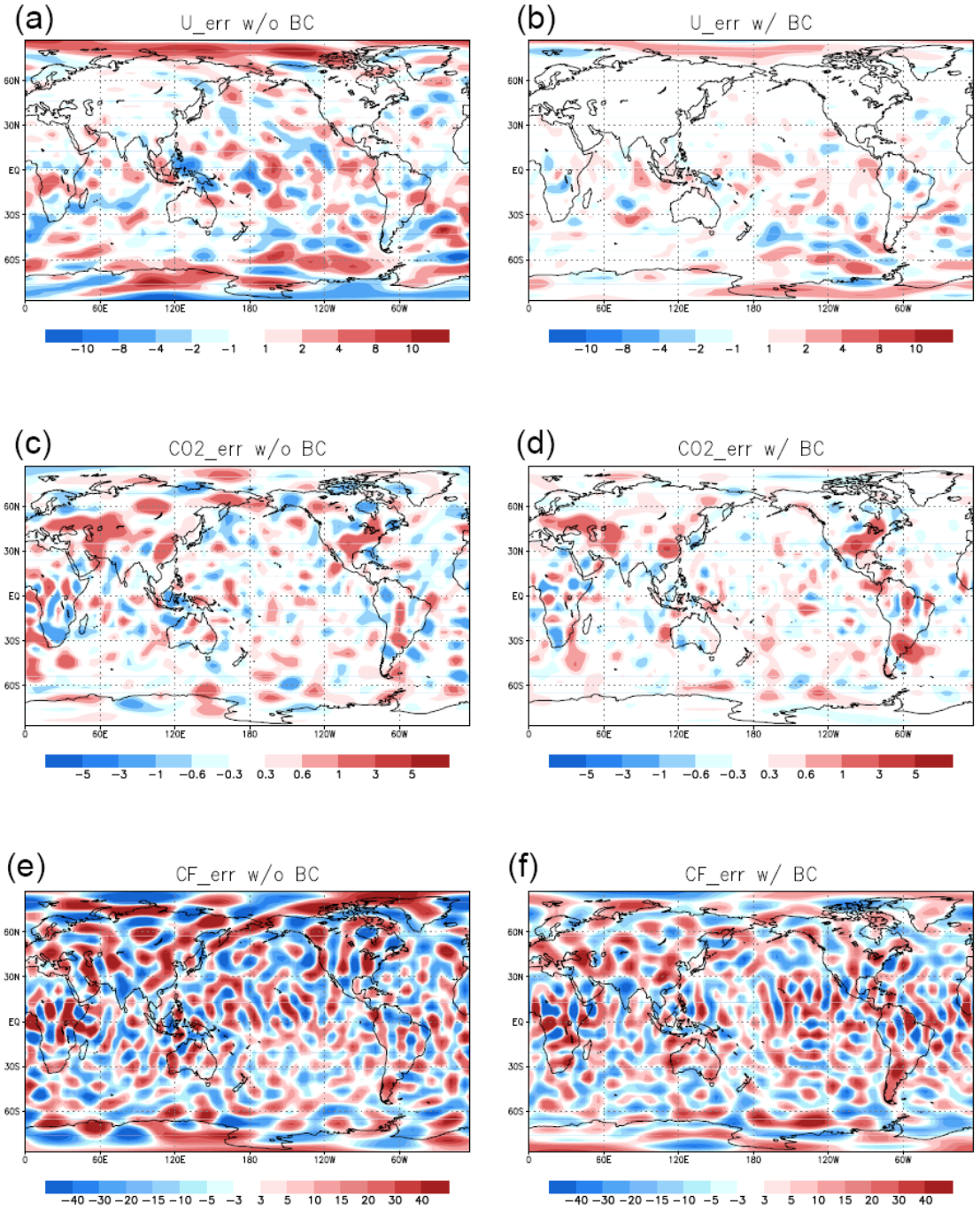

Figure 4. 6. The spatial distribution of errors in the analysis (analysis minus truth): zonal wind (m/s) (a) without bias correction, (b) with bias correction, atmospheric CO_2 (ppmv) on the bottom layer (c) without bias correction, (d) with bias correction, and surface CO_2 fluxes (10^{-9} kg/m^2/s) (e) without bias correction, and (f) with bias correction, after two month of data assimilation.

indicated by the RMS errors (Figure 4.5). Since the atmospheric CO_2 is observed every other grid point, the analysis without bias correction is not too bad in general, but the bias correction of wind fields obviously improves the analysis of atmospheric CO_2 fields. In the Figures 4.6 (e) and (f), the global maps of surface CO_2 flux analysis have very noisy errors all over the region even after the bias correction although the spatial phases of positive and negative signals seem generally matched with true state and the strength of these noisy signals is slightly weaker.

4.4. Adaptive inflation

So far, a multiplicative covariance inflation of 8% (obtained by tuning the inflation parameter in the atmospheric analysis system) has been used in order to prevent the analysis system from underestimating the background error covariance, which is fixed in time. With a constant multiplicative covariance inflation, the bias correction made a significant improvement on meteorological variables in Section 4.3. However, the improvement in the CO_2 analysis was only marginal in the imperfect model experiment. Now, we implement an adaptive inflation technique which can reflect the observations efficiently according to the quality of the analysis and the background in time. Also, we deal with the inflation of the atmospheric CO_2 separately from that of other meteorological variables, in addition to using a different inflation on each vertical layer. The adaptive inflation technique allows us to avoid tuning these inflation factors, since it would be very expensive or even infeasible to find reasonable levels of all these inflation factors by manually tuning.

4.4.1. Methodology for variable having observation

The basic idea on the adaptive inflation comes from LI09. Let's assume that the background error covariance, $\mathbf{P}^b$, and the observation error covariance, $\mathbf{R}$, are correctly specified. If the errors of the background and the observations are not correlated, then one can write an equation that relates the observational increments (that we can measure) and the error covariances as follows:

$$< \mathbf{d}_{o-b}\mathbf{d}_{o-b}^T >= \mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R} \qquad (4.1)$$

where $\mathbf{d}_{o-b}$ is the difference between observations and the corresponding background at the observation space and the brackets indicate an average over many cases. However, we know that it is necessary to inflate the $\mathbf{P}^b$ since it tends to be underestimated in practice. Thus, $\mathbf{P}^b$ in Equation (4.1) should be multiplied by the inflation factor, Δ (larger than unity). From this, and transposing (4.1) the adaptive inflation can be obtained as

$$\Delta = \frac{\mathbf{d}_{o-b}^T\mathbf{d}_{o-b} - Tr(\mathbf{R})}{Tr(\mathbf{H}\mathbf{P}^b\mathbf{H}^T)} \qquad (4.2)$$

However, we cannot use this equation to estimate Δ because (4.1) or (4.2) are based on the assumption that the observation error covariance is accurate, something not true in practice. If $\mathbf{R}$ is not known precisely, the use of (4.2) fails because its errors are compensated by the value of Δ.

Another diagnostic on background errors comes from the combination of $\mathbf{d}_{o-b}$ and *analysis-minus-background* $\mathbf{d}_{a-b}$ (Desroziers et al., 2005). From this method, the relationship

$$< \mathbf{d}_{a-b}\mathbf{d}_{o-b}^T >= \mathbf{H}\mathbf{P}^b\mathbf{H}^T \qquad (4.3)$$

has been derived by Desroziers et al. (2005), and Equation (4.3) produces another alternative formulation of an estimate of the inflation:

$$\Delta = \frac{\mathbf{d}_{a-b}^{T}\mathbf{d}_{o-b}}{Tr(\mathbf{H}\mathbf{P}^{b}\mathbf{H}^{T})} \qquad (4.4)$$

after multiplying the background error covariance matrix by the inflation factor and transposing as before. Equation (4.2) and (4.4) will be referred to as OMB^2 and AMB*OMB estimations, respectively.

In addition, LI09 pointed out that it is necessary to have a correct observation error covariance, $\mathbf{R}$, for an accurate estimate of adaptive inflation from those methods. Thus, LI09 calculates the observation errors simultaneously from the relationship for $\mathbf{R}$ proven by Desroziers et al. (2005):

$$< \mathbf{d}_{o-a}\mathbf{d}_{o-b}^{T} >= \mathbf{R} \qquad (4.5)$$

where $\mathbf{d}_{o-a}$ ($\mathbf{d}_{o-b}$) are the difference between the observation and analysis (background) in observation space. Taking the transpose of Equation (4.5) allows estimating the observation errors as follows:

$$(\tilde{\sigma}_o^2)_i = (\mathbf{d}_{o-a})^T(\mathbf{d}_{o-b})/\,p_i = \sum_{j=1}^{p_i}(y_j^o - y_j^a)(y_j^o - y_j^b)/\,p_i \qquad (4.6)$$

where y_j^o is the value of observation j and y_j^a, y_j^b are their analysis and background at the observation space j. Equation (4.6) calculates the variance of any subset of observations i with p_i observations, and LI09 called OMA*OMB estimation.

LI09 estimated the adaptive inflation factor and observation error variance simultaneously with temporal smoothing (Kalnay 2003, Appendix C) in order to reduce the problem of sampling error. That is, one can tune a "forgetting" parameter

$\kappa > 1.0$ which inhibits temporally drastic changes in the estimated adaptive inflation and observation errors. Thus, the final estimates are determined by a balance between the estimates at the previous and the current analysis time steps according to the magnitude of κ. Since the ratio of the weight for the current estimate to that for the previous one is κ -1, the resultant estimates of adaptive inflation and observation errors forget the previous estimates less as κ decreases. LI09 mentioned that the final estimate is not sensitive to this forgetting parameter, and our work also confirmed that.

While LI09 tested for the atmospheric variables and estimated both inflation and observation errors assuming one value of inflation for all model grids, we applied this technique to our carbon cycle data assimilation system and estimated the adaptive inflation and observation errors on each vertical layer separately. In addition, for the atmospheric CO_2 on the lowest layer, the inflation over the land and the ocean were estimated separately. That is because the observation network of wind fields is dense over the land (Figure 2.3) so that it tends to suppress the ensemble spread of not only the wind field but also the atmospheric CO_2 over land (Figure 4.7) if we use the same inflation for all grid points at the lowest layer's atmospheric CO_2. The spread in Figure 4.7 represents the value after multiplying the inflation factor to the background ensemble spread, which is used directly for the analysis cycle, in a test experiment using a single value of inflation for the atmospheric CO_2 on the lowest layer. Figure 4.7(c) indicates that there is a lack of ensemble spread over the land in atmospheric CO_2 background. This result can be explained as follows: atmospheric CO_2 is transported by the wind fields so that the improved wind analysis reduces the

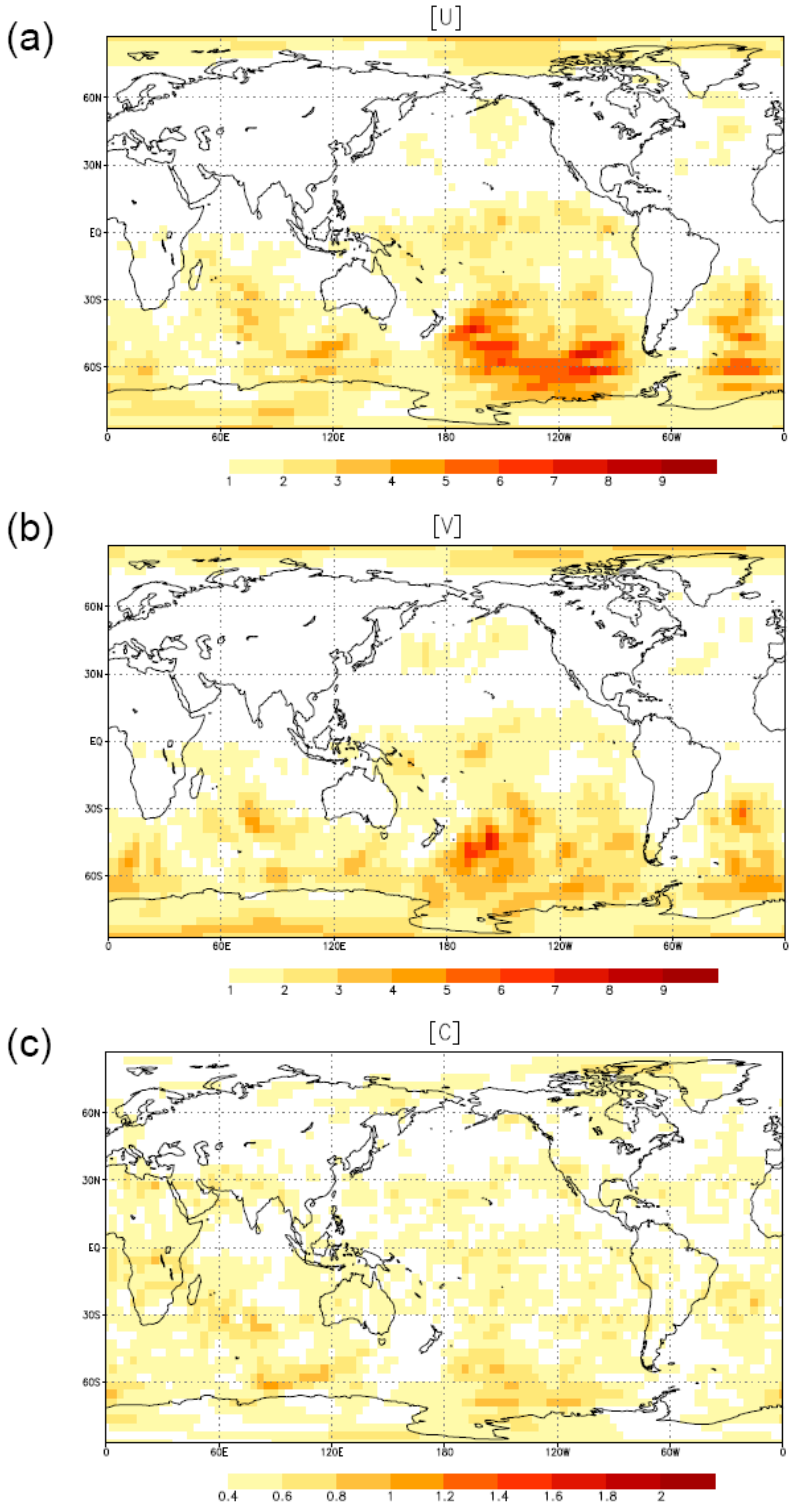

Figure 4. 7. Spread of background ensemble in (a) zonal wind (U), (b) meridional wind (V), and (c) atmospheric CO_2 concentration after three weeks of data assimilation under the experiment using a single inflation for the atmospheric CO_2 at the lowest layer.

uncertainty of atmospheric CO_2. However, the wind observations are located mainly over land. Therefore, the analysis cycle of one-way multivariate data assimilation reduces the ensemble spread of atmospheric CO_2 analysis more over land than over the oceans when we use the horizontal average of adaptive inflation. However, CO_2 observations, which have a uniform distribution in the horizontal, should be reflected in the analysis not only over the ocean but also over the land. That is because the lowest layer's CO_2 concentration is directly related to the surface CO_2 fluxes and we have to resolve the dominant variation of surface CO_2 fluxes over land as precisely as possible.

4.4.2. Adaptive inflation for a variable having no observation

Now, we need to take account of the inflation factor for a variable which has no observation such as the surface CO_2 fluxes in our case. That is because the method prescribed in Section 4.4.1 is based on measuring observational increments, and therefore is only valid for variables that are observed, but one important component of state vector in this study, namely surface CO_2 fluxes, does not have observations (at least no remotely sensed observations).

We conducted first several sensitivity experiments with different fixed inflations for the surface CO_2 fluxes. We found that a relatively large inflation, compared to the inflation for other dynamic variables of analysis, made the analysis diverge and only smaller values such as less than 5% kept the analysis stable. We now give a mathematical argument in support of this experimental result, namely that unobserved variables should have a smaller inflation than observed variables.

Consider the simplest case of a state vector with just two components: one observed (x_{b1}), the other not observed (x_{b2}). State vector $\mathbf{x}_b$ includes x_{b1} and x_{b2}, and the observation operator H should be H=(1 0) so that $\mathbf{H}\mathbf{x}_b = y_{b1} : y_o = y_{o1}$ since x_{b2} is not observed. Let $\Delta = (\Delta_1 \quad \Delta_2)^T$ be the vector of inflation factors (>1), then $\left\langle (y_{o1} - y_{b1})^2 \right\rangle = \sigma_{o1}^2 + \Delta_1 \sigma_{b1}^2$. We could calculate the inflation factor, Δ_1 , and observation error, σ_{o1}^2, by the method in Section 4.4.1, but the equations do not give any information about x_{b2} and Δ_2.

Within the LETKF,

$\mathbf{x_a} - \mathbf{x_b} = \mathbf{K}(\mathbf{y_o} - \mathbf{H}\mathbf{x_b})$ where $\mathbf{K} = \mathbf{P}^b\mathbf{H}^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}$ and

$$\mathbf{P}^b = \begin{pmatrix} \varepsilon_{b1}^2 & \varepsilon_{b1}\varepsilon_{b2} \\ \varepsilon_{b1}\varepsilon_{b2} & \varepsilon_{b2}^2 \end{pmatrix} \qquad (4.7)$$

Then, $\mathbf{P}^b\mathbf{H}^T = \begin{pmatrix} \varepsilon_{b1}^2 \\ \varepsilon_{b1}\varepsilon_{b2} \end{pmatrix}$ and $\mathbf{K} = \dfrac{\begin{bmatrix} \left(\sqrt{\Delta_1}\varepsilon_{b1}\right)^2 \\ \sqrt{\Delta_1}\varepsilon_{b1}\sqrt{\Delta_2}\varepsilon_{b2} \end{bmatrix}}{\sigma_{o1}^2 + \left(\sqrt{\Delta_1}\varepsilon_{b1}\right)^2}$. Assume that σ_{o1}^2 is perfectly known and small compared to $\left(\sqrt{\Delta_1}\varepsilon_{b1}\right)^2$, then we can approximate $y_{o1} - Hx_{b1} = (y_{o1} - y_{b1}) \cong -\varepsilon_{b1}$. From the equation (4.7),

$$\begin{pmatrix} x_{a1} - x_{b1} \\ x_{a2} - x_{b2} \end{pmatrix} = \begin{pmatrix} \left(\sqrt{\Delta_1}\varepsilon_{b1}\right)^2 / \left[\sigma_{o1}^2 + \left(\sqrt{\Delta_1}\varepsilon_{b1}\right)^2\right] \\ \sqrt{\Delta_1}\varepsilon_{b1}\sqrt{\Delta_2}\varepsilon_{b2} / \left[\sigma_{o1}^2 + \left(\sqrt{\Delta_1}\varepsilon_{b1}\right)^2\right] \end{pmatrix} (y_{o1} - y_{b1}) \qquad (4.8)$$

If the filtering is working properly, we can also assume that $x_{a2} - x_{b2} = \varepsilon_{a2} - \varepsilon_{b2} \cong -\delta\varepsilon_{b2}$ with $\delta < 1$. Then,

$$x_{a2} - x_{b2} = -\delta\varepsilon_{b2} = \left\{\sqrt{\Delta_1}\varepsilon_{b1}\sqrt{\Delta_2}\varepsilon_{b2} \Big/ \left[\sigma_{o1}^2 + \left(\sqrt{\Delta_1}\varepsilon_{b1}\right)^2\right]\right\}\left(-\varepsilon_{b1}\right).$$

According to the assumption of $\sigma_{o1}^2 << \left(\sqrt{\Delta_1}\varepsilon_{b1}\right)^2$, $\delta\varepsilon_{b2} = \left\{\sqrt{\Delta_1}\varepsilon_{b1}\sqrt{\Delta_2}\varepsilon_{b2} \Big/ \left[\left(\sqrt{\Delta_1}\varepsilon_{b1}\right)^2\right]\right\}\varepsilon_{b1}$. From this equation, we find that $\frac{\sqrt{\Delta_2}}{\sqrt{\Delta_1}} \cong \delta < 1$, that is, the inflation for the variable having no observation should be smaller than that for the variables having observations. Although this derivation is proven when $\sigma_{o1}^2 << \left(\sqrt{\Delta_1}\varepsilon_{b1}\right)^2$, in reality, this assumption may not be always true. But, empirically, we have experienced a failure of surface CO_2 flux analysis in the imperfect model experiments when using a similar or larger inflation factor for the surface CO_2 flux compared with those for the other observed variables. This empirical evidence supports the result of the derivation.

So far, we have proven that the inflation for the surface CO_2 fluxes should be less than that for other variables having observation. Then, the issue is how to adaptively estimate the inflation for a variable which does not have an observation. Here, what we tried is basically to let the global ensemble spread of the analysis be the same as that of the forecast. The temporal smoothing introduced in the previous Section 4.4.1 has been also applied with the same value of the forgetting parameter. The formulation is similar to the inflation of covariance relaxation method (Zhang et al., 2004). We calculate the inflation for every grid point at every analysis time.

Then, we will see the magnitude of estimated inflation and whether it will be less than the inflation of other variables having observations.

4.4.3. Results from the adaptive inflation technique

First, it is essential to calculate the observation errors accurately in order to have reasonable estimates of the adaptive inflations. Thus, we first checked the observation error estimate from the OMA*OMB method. Tables 4.1 and 4.2 show that the online estimate of observation error at every vertical layer has a good convergence to the true value, although the initial guess of each is set at a double of the true value. Furthermore, the estimates converged to the reasonable range of the true value very quickly, only after several days of analysis.

By estimating adaptive inflation as explained in Sections 4.4.1 and 4.4.2 in addition to the bias correction, we could get a very good analysis of surface CO_2 fluxes as well as an improvement in the analysis of other variables. As in the results of LI09, using OMB^2 or AMB*OMB to estimate the inflation also have similar performance. In the RMS error plot of Figure 4.8, it is possible to avoid filter divergence in the analysis of surface CO_2 fluxes after implementing adaptive inflation estimation. Owing to the stable analysis of surface flux forcing, the analysis of atmospheric CO_2 concentration fields has been improved as well. The spatial distribution of surface CO_2 fluxes compared to the true state (Figure 4.9) is vastly improved compared to the case without adaptive inflation and only bias correction. Figure 4.10 represents the improvement from the adaptive inflation and observation error estimation in terms of RMS errors for all variables and several regions. From

Table 4. 1. Estimated observation error standard deviations, using the OMB^2 method (results after two months of analysis).

Vari-able	True value	Initial Guess	Estimated observation error for each vertical layer						
			level 1	level 2	level 3	level 4	level 5	level 6	level 7
U	1	2	1.076	1.111	1.091	1.168	1.221	1.208	1.026
V	1	2	1.042	1.108	1.117	1.159	1.220	1.192	1.055
T	1	2	1.059	1.017	1.059	1.056	1.075	1.068	1.008
q	0.1	0.2	0.272	0.165	0.120	0.114	0.112	0.102	0.100
CO2	1	2	1.000	0.958	0.962	0.973	0.964	0.974	0.972
Ps	100	200	105.701						

Table 4. 2. Estimated observation error standard deviations, using the AMB*OMB method. (results after two months of analysis)

Vari-able	True value	Initial Guess	Estimated observation error for each vertical layer						
			level 1	level 2	level 3	level 4	level 5	level 6	level 7
U	1	2	1.135	1.132	1.131	1.202	1.272	1.230	1.031
V	1	2	1.072	1.142	1.126	1.173	1.211	1.171	1.065
T	1	2	1.068	1.036	1.051	1.077	1.087	1.047	1.031
q	0.1	0.2	0.302	0.197	0.128	0.120	0.112	0.101	0.101
CO2	1	2	1.000	1.012	1.011	1.018	0.994	1.008	1.005
Ps	100	200	111.352						

the chart, one can say that the impact of adaptive inflation and observation error estimate is huge in the analysis of CO_2 surface fluxes. Since the contamination from the poor analysis of surface CO_2 fluxes is removed by the adaptive inflation, the atmospheric CO_2 analysis gets improved (Figures 4.8 and 4.10). Also, it improves the analysis of all atmospheric variables except for specific humidity. The quality of the humidity analysis is not improved as much as the other variables and this is partially attributed to the observation error estimate for the humidity. Tables 4.1 and

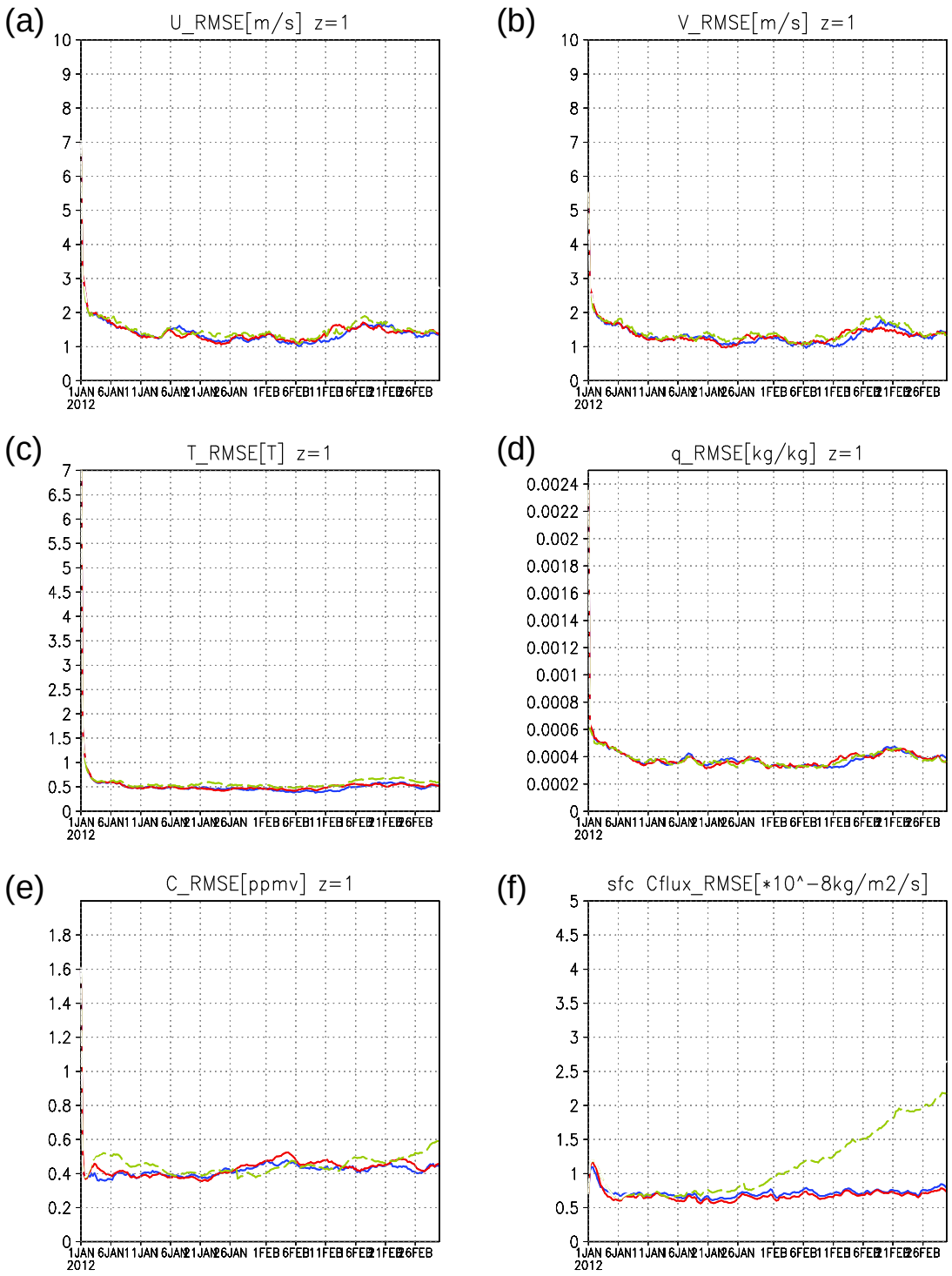

Figure 4. 8. RMS errors of (a) U, (b) V, (c) T, (d) q, (e) atmospheric CO_2 at the level of σ=0.95, and (f) surface CO_2 fluxes in the analysis. (blue: with bias correction and adaptive inflation with the OMB^2 method, red: with bias correction and adaptive inflation with the AMB*OMB method, green: with bias correction, but no adaptive inflation)

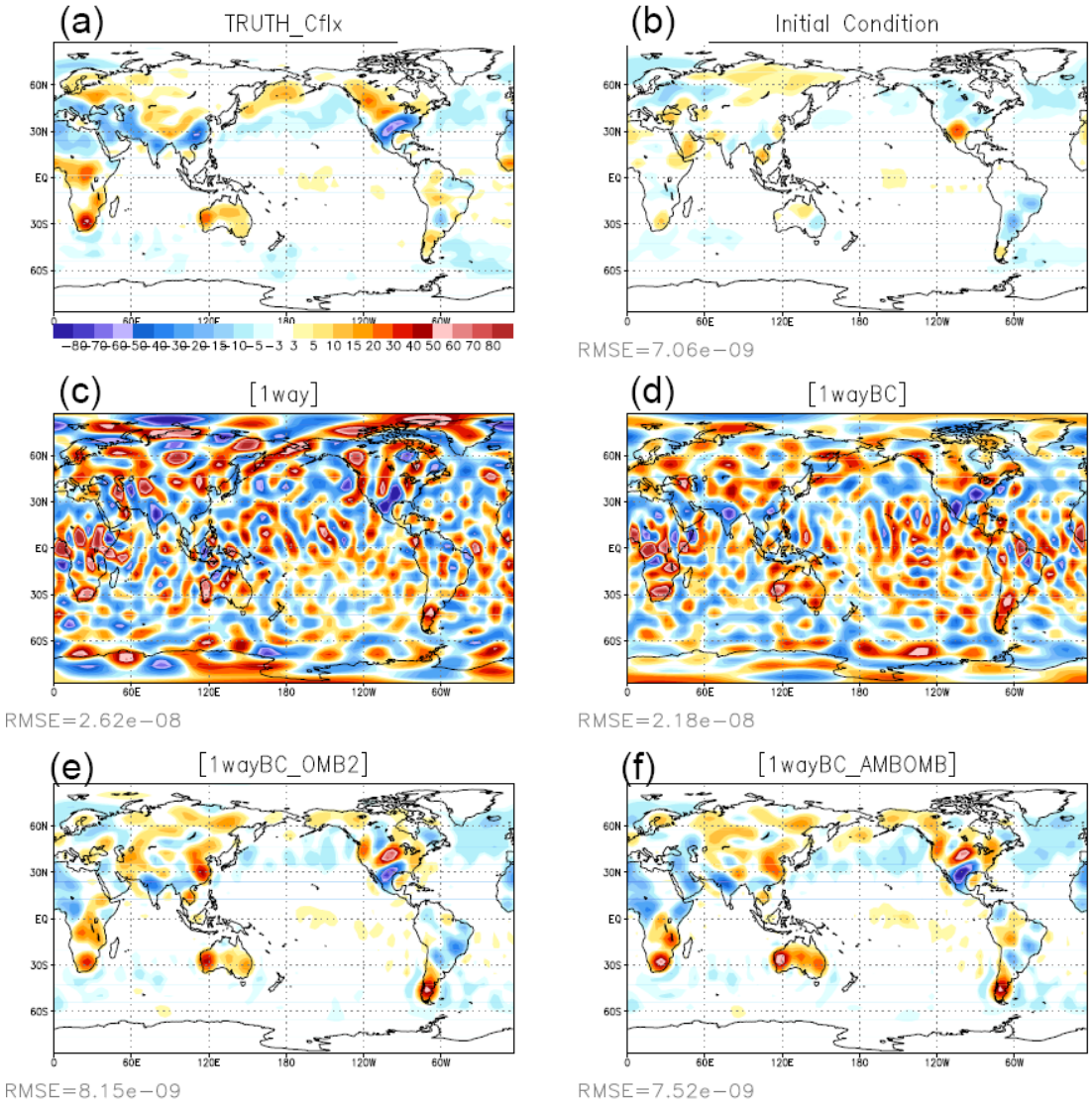

Figure 4. 9. Same as Figure 4.8, except for surface CO_2 flux fields. (unit: 10^{-9} kg/m^2/s)

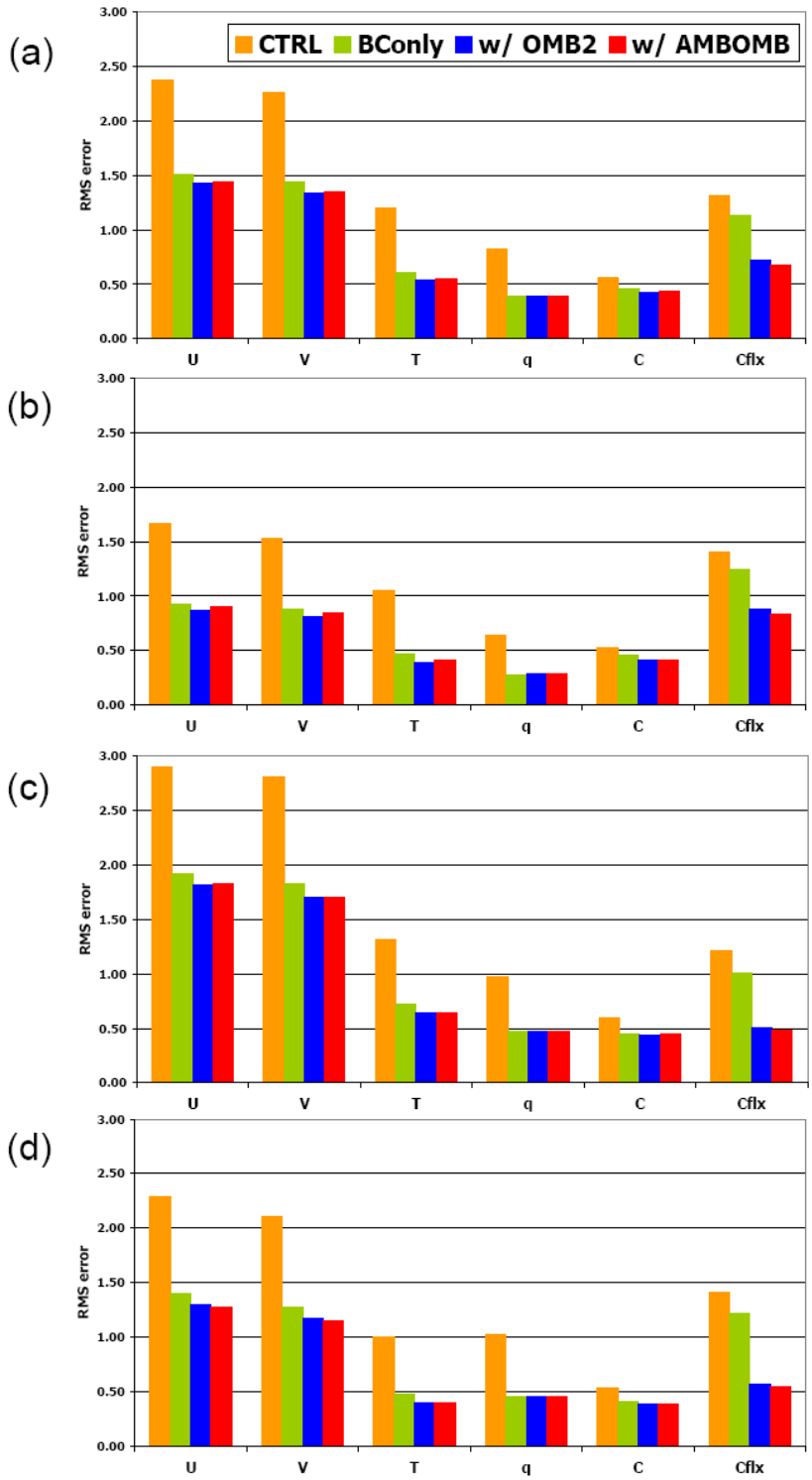

Figure 4. 10. RMS error for two months of analysis: (a) global total, (b) Northern Hemisphere, (c) Southern Hemisphere, and (d) Tropics (20°S ~ 20°N). Yellow bar indicates CTRL experiment, green bar results from the bias correction experiment, blue bar is from the experiment of the bias correction plus adaptive inflation of the OMB^2 method, and red bar presents the result of the bias correction plus adaptive inflation of the AMB*OMB method. (unit: m/s for U and V, K for T, g/kg for q, ppmv for atmospheric CO_2 (C), $10^{-8}kg/m^2/s$ for surface CO_2 fluxes (Cflx))

4.2 show that the estimate of observation error for q is not accurate enough on the bottom two layers (Table 4.1 and 4.2), but it does not degrade it either.

For the surface CO_2 flux analysis, the adaptive inflation technique made the crucial impact eliminating spurious noises as well as helping produce reasonable results under the imperfect model assumption which does not include any priori information and direct observation. In detail, we could see the AMB*OMB method has slightly better results in the analysis of surface CO_2 fluxes in terms of RMS error and spatial distribution. Analysis from the AMB*OMB method is closer to the true state especially over the Southern Africa, South America, and Asia compared to the OMB^2 method (Figure 4.9).

Even though the final analyses from both AMB*OMB method and OMB^2 method have similar performances, we found that their estimated inflations are somewhat different from one another. Figure 4.11 has the estimated inflation factors from OMB^2 method and Figure 4.12 from AMB*OMB method. In general, OMB^2 method has larger magnitude of estimated inflations and the values are varying with time more drastically than AMB*OMB. The reason why the OMB^2 method has more noisy patterns in the time series of resultant inflation parameters (Figure 4.11) is because the term of Tr(**R**), tracer of observation error covariance matrix, is subtracted explicitly in the formulation. Every analysis step calculates the observation error statistics and uses them for estimating adaptive inflation explicitly in OMB2 method. Although the estimates of observation error have very good agreement to the standard deviation of true observation errors through OMA*OMB method, actual deviation of

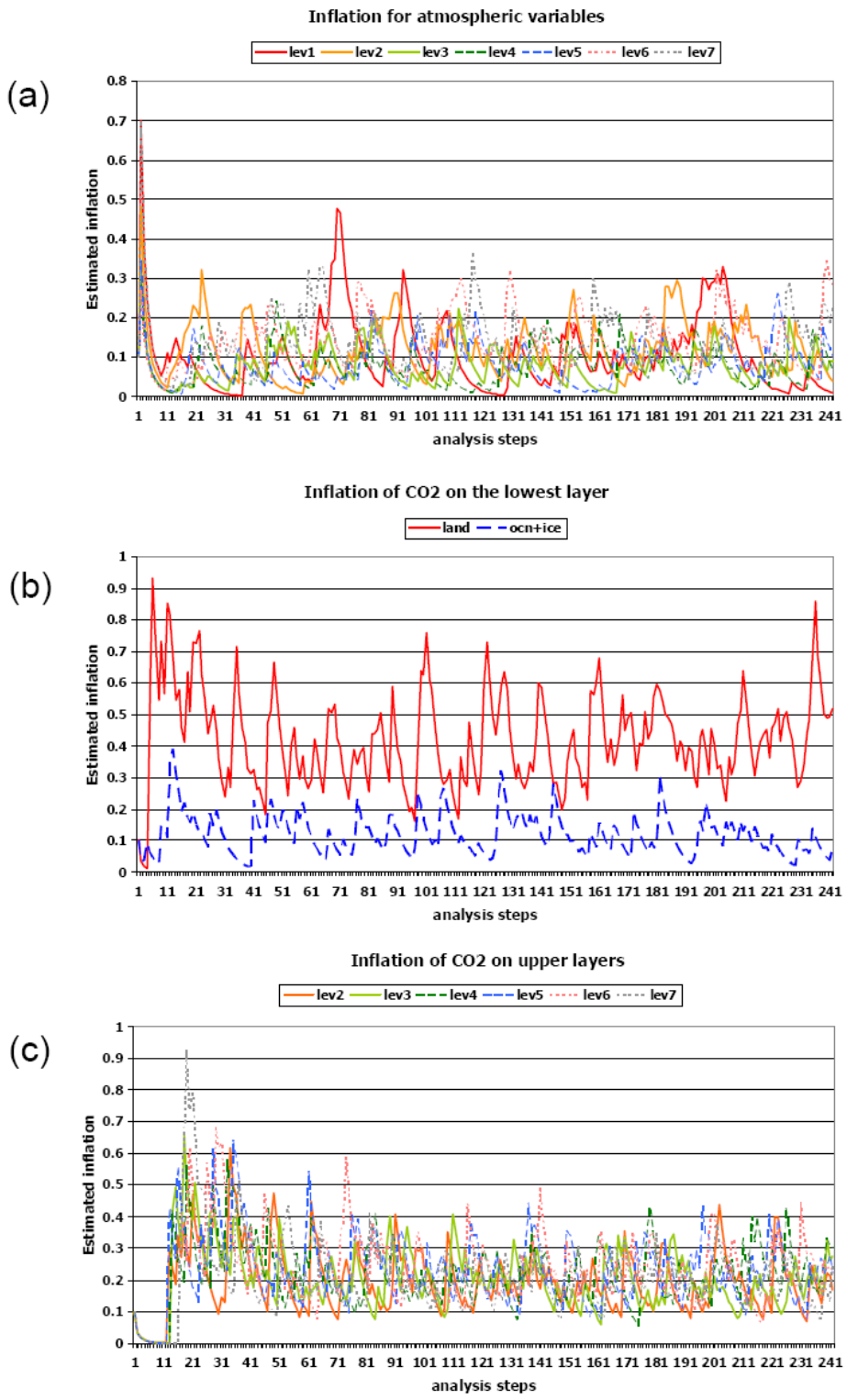

Figure 4. 11. Time series of resultant adaptive inflations ($\Delta - 1$) through the OMB^2 method for (a) meteological variables for all vertical levels, (b) atmospheric CO_2 on the bottom layer (red: over land, blue: over ocean), and (c) atmospheric CO_2 on upper levels

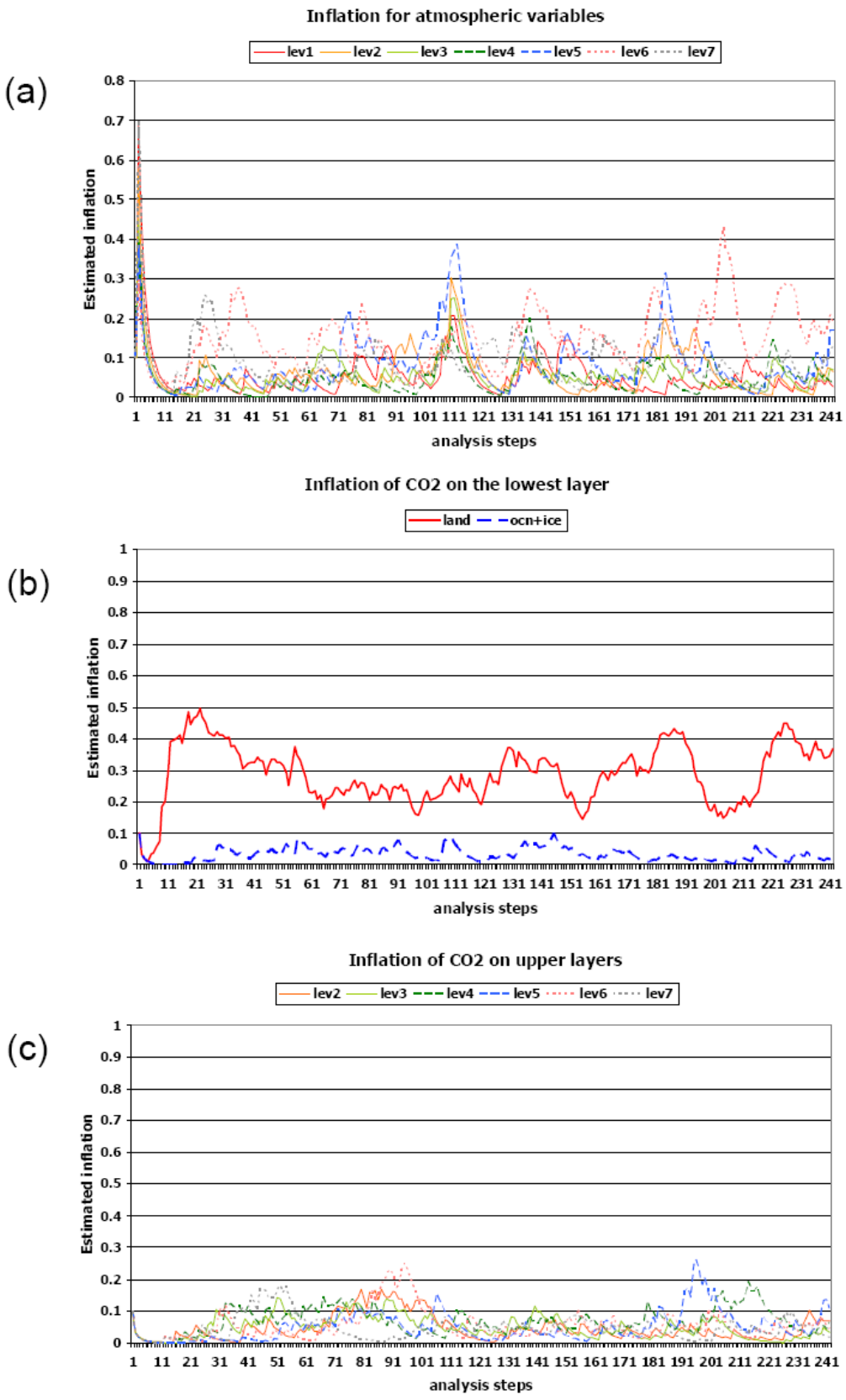

Figure 4. 12. Same as Figure 4.11, except for AMB*OMB method.

observation error estimate from the real observation error cannot be sometimes negligible and the differences are fluctuated with time. Thus, the resultant inflation varies more strongly depending on the estimated observation errors. We note that if the estimated adaptive inflation has a negative value in both methods, we set the value as zero and then apply the temporal smoothing so that there is never negative inflation applied to the background.

For the inflation of atmospheric CO_2 on the lowest layer, both methods produced larger values over the land than over the ocean. This result is consistent with what we expected. Because of larger inflation of the atmospheric CO_2 over the land, the analysis can reflect the observation more over the land where the variation of surface CO_2 fluxes is dominant in the nature. The successful performance of adaptive inflation is attributed to the reasonable estimate of observation errors because the calculation of adaptive inflation counts on the accurate observation error estimate. LI09 showed that we cannot have good results from the adaptive inflation if the estimation of observation error is not correct.

Moreover, the adaptive inflation for surface CO_2 fluxes has also generated a reasonable value overall. Figure 4.13 displays the average of adaptive inflation over the analysis period and it turns out that the area where the variation of surface CO_2 fluxes are relatively large has more inflation over the land while the inflation over ocean has relatively small values. Furthermore, the magnitude of inflation for the surface CO_2 fluxes is generally less than that for other variables having observations, compared to Figure 4.11 and 4.12. This is physically and quantitatively reasonable

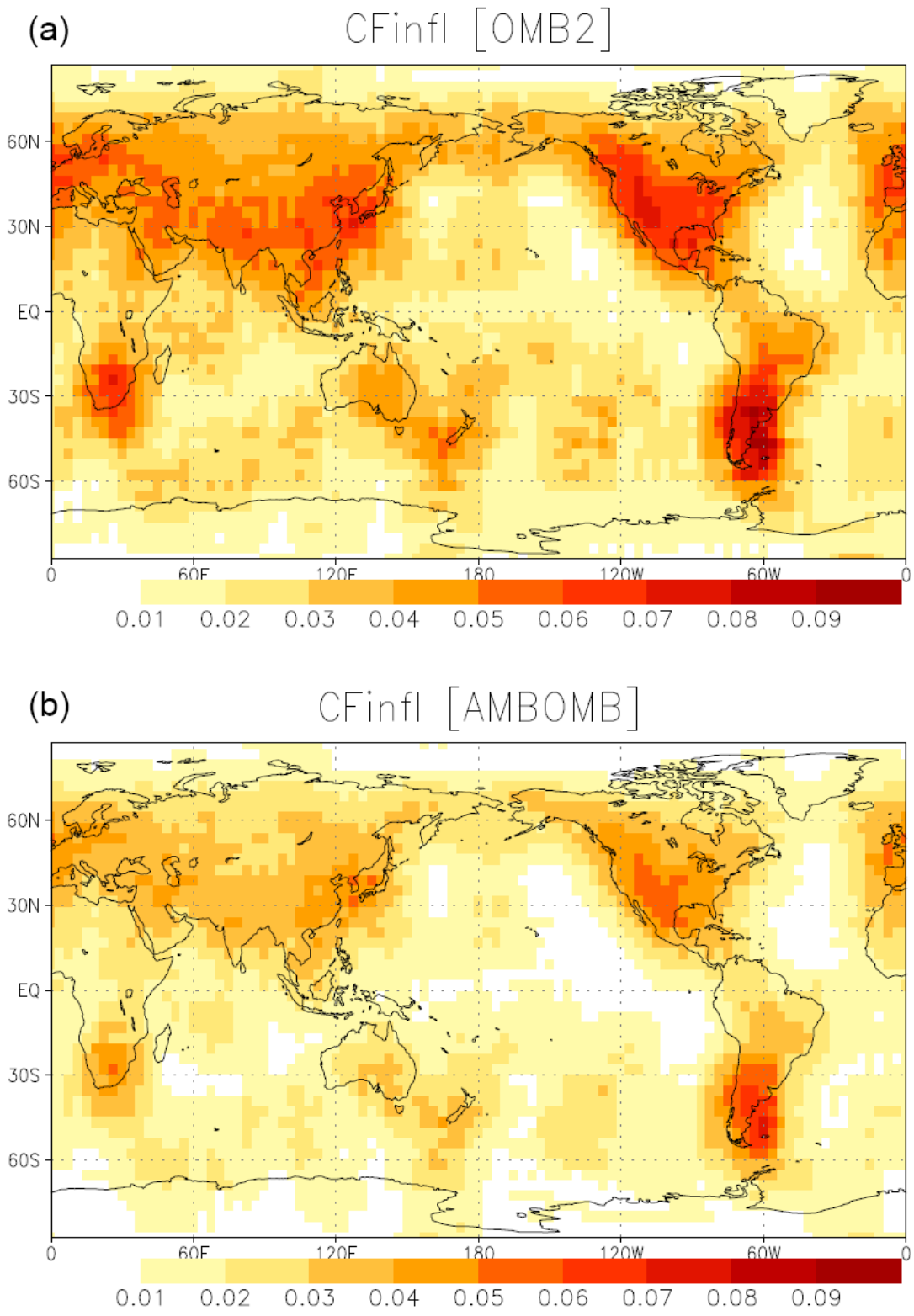

Figure 4. 13. Resultant adaptive inflation ($\Delta - 1$) for surface CO_2 fluxes for two months of analysis period, coupled with (a) the OMB^2 method, and (b) the AMB*OMB method.

based on our discussion in Section 4.4.2 suggesting that the adaptive inflation technique of this variable works properly.

4.5. Summary

One-way multivariate data assimilation has been tested under an imperfect model assumption. Since the climatology of the nature run is significantly different from that of the forecast model, the ensemble Kalman filter cannot represent the analysis properly. Thus, we implemented the low-dimensional bias correction of Danforth et al. (2007) and got a significantly improved analysis of atmospheric variables and atmospheric CO_2. Even with a good analysis of these variables, however, we could not meet an acceptable range of accuracy in surface CO_2 flux analysis which was too noisy.

After several sensitivity experiments with different inflation factors for surface CO_2 fluxes, we found that using a small inflation for surface carbon helped the analysis not to diverge; and then we proved mathematically that inflation for unobserved variables should be smaller than that for the observed. This brought us to think about adaptive inflation estimation. We examined the advanced technique of a simultaneous estimation of adaptive inflation and observation error introduced by LI09 for the variables having observations. In the mean time, we also investigated another adaptive inflation method for surface CO_2 fluxes which are never observed in our experiments. With these adaptive inflation methods, we could get much more stable and reasonable analysis of surface CO_2 fluxes. As our mathematical derivation suggested, we found that inflation for surface CO_2 fluxes should be smaller than for

other variables, while the inflation of atmospheric CO_2 tends to be larger than that of other atmospheric variables, especially near the surface. Moreover, the inflation of atmospheric CO_2 is estimated to be relatively large at the beginning of the analysis period because the random initial condition of surface CO_2 flux causes an incorrect forecast of atmospheric CO_2 so that analysis system should consider the observation more during the first several days. In a similar manner, the inflation of the surface layer's atmospheric CO_2 is larger over the land than over the ocean. That is because a variability of surface CO_2 fluxes is dominant over land and it is reflected in the observed atmospheric CO_2 in the lowest layer. Thus, the analysis system should be sensitive to the CO_2 observation over the land more than over the ocean in order to estimate surface CO_2 fluxes precisely.

Lastly, we also calculated the observation error through LI09's method and the accuracy of estimated observation error was excellent, with the analysis converging close to the true value even though the initial guess of it was set to be double the true value for each variable. We should point out that this correct estimation of observation error is essential in making the adaptive inflation work well enough for the CO_2 analyses, and most importantly, it should be extremely useful with real CO_2 data assimilation, where the estimates of observation errors will be very uncertain.

Chapter 5: Application of Adaptive Inflation and Estimation of Observation Errors to the Perfect Model Simulation

5.1. Introduction

Recall that in Chapter 2 we found that the C-univariate data assimilation with a fixed inflation factor had diverging CO_2 analysis, and that increasing the inflation only increased the rate of divergence. In Chapter 4, we found that for an imperfect model, the use of model bias estimation and adaptive inflation/observation errors was able to recover a very reasonable estimation of surface carbon fluxes. We therefore decided to apply the technique of adaptive inflation and observation error estimation to the perfect model simulation to check whether the adaptive inflation can avoid filter divergence in the C-univariate data assimilation work and how much it impacts both C-univariate and one-way multivariate data assimilation for CO_2 analysis. We would also like to see whether the other atmospheric variables are also improved by implementing adaptive inflation.

5.2. Experimental Design

The basic setting of the experiments is same as that of chapter 2 (section 2.4.1) in terms of the initial conditions and observations except for estimating the adaptive inflation and observation errors. Moreover, only two types of analysis,

carbon-univariate data assimilation and 1-way multivariate data assimilation, are examined in this chapter. The initial guesses of observation errors have been given as twice the true values, and the inflation starts from 10% at the initial time. Since the OMB^2 method had a comparable performance with the AMB*OMB for the CO_2 analysis, we show the result from the OMB^2 method estimating adaptive inflation coupled with the OMA*OMB method for simultaneous estimation of observation errors here. For the adaptive inflation of the surface CO_2 fluxes, we applied the same techniques described in Section 4.3.3. Since this is a simple case, we do not use different inflation in the horizontal even for the atmospheric CO_2 and surface CO_2 fluxes. That is, one constant of adaptive inflation is estimated for each vertical layer and for each variable.

5.3. Results

For the atmospheric variables, the improvement brought by adaptive inflation is not significant enough to be visually apparent (Figure 5.1 and Figure 2.5). This is somewhat different from LI09's result in which the atmospheric analysis had significant improvements with the same method. This is because LI09 assumed atmospheric observations at every other grid point whereas our observation network has much less density (rawinsonde distribution, Figure 2.3). While there is no remarkable improvement of the atmospheric analysis, the system with adaptive inflation and observation error estimation is working properly and has major impacts on the analysis of CO_2 variables. Indeed, adaptive inflation and observation error

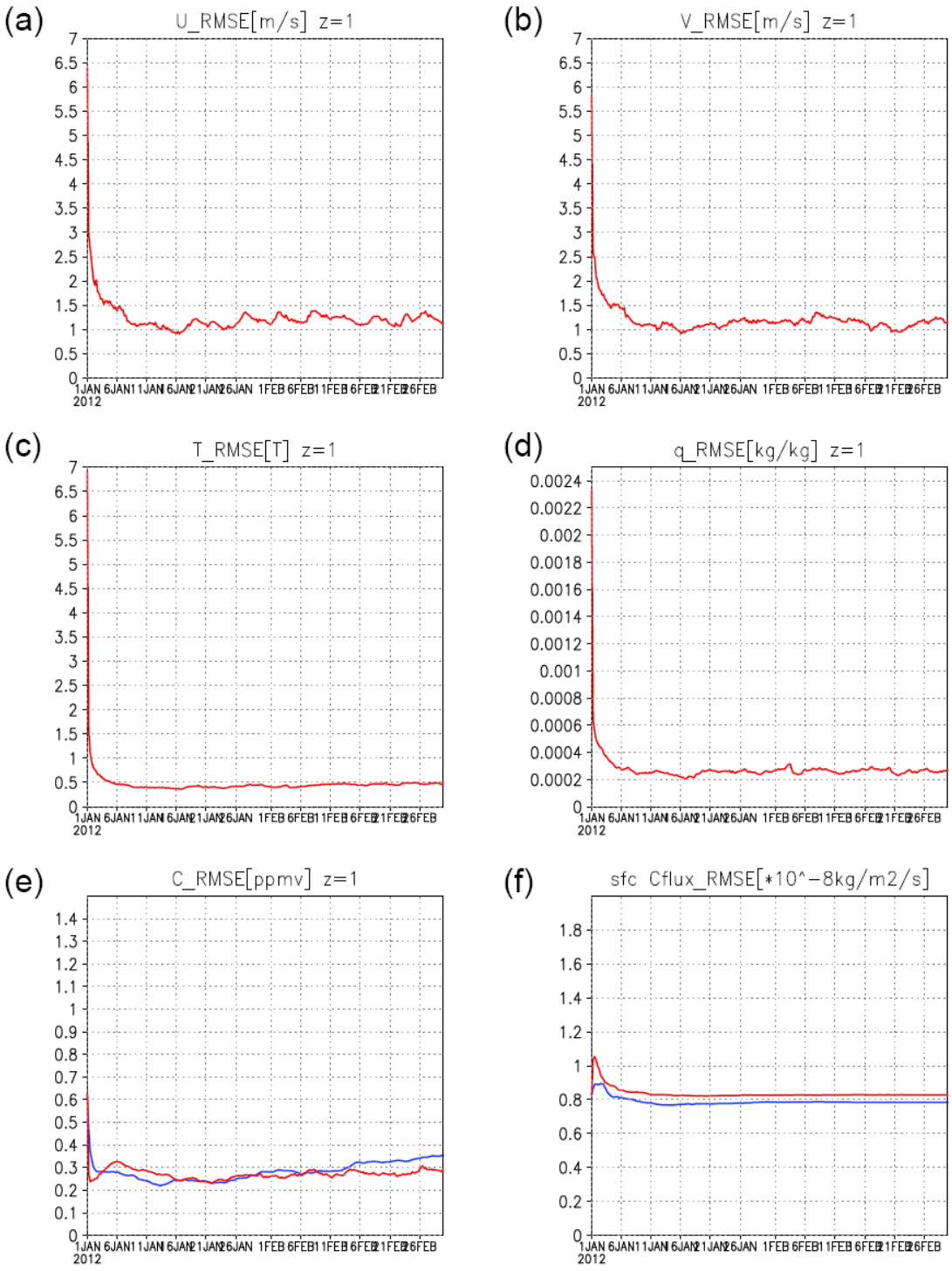

Figure 5. 1. RMS errors of (a) U (m/s), (b) V (m/s), (c) T (K), (d) q (kg/kg), (e) atmospheric CO_2 concentration (ppmv) at the level of σ=0.95, and (f) surface CO_2 fluxes (*10-8kg/m2/s). (blue: C-univariate analysis with adaptive inflation, red: one-way multivariate analysis with adaptive inflation)

estimates makes significant improvement on the CO_2 analysis of the C-univariate data assimilation, which had diverged with fixed inflation for the CO_2 variables (blue lines in Figure 5.1 and green lines in Figure 2.5). On the other hand, 1-way multivariate data assimilation also has better results, but not as remarkable as the improvement on the carbon-univariate data assimilation.

Figures 5.2 and 5.3 show a comparison of 1-way multivariate data assimilation and C-univariate data assimilation with and without the adaptive inflation technique in the global maps of atmospheric CO_2 on the lowest layer and surface CO_2 fluxes after two months of analysis. First, the atmospheric CO_2 fields (Figure 5.2) show that 1-way multivariate analysis still has better result than the C-univariate in terms of both RMS error and spatial distribution. In the RMS error, it is 0.297 ppmv in 1-way multivariate data assimilation with adaptive inflation and 0.325 ppmv in the C-univariate, but one can say both are good enough considering the observation error of atmospheric CO_2 concentration is 1.0 ppmv. This technique has a huge impact on the C-univariate data assimilation which diverged due to a failure of analyzing surface CO_2 fluxes noted in Chapter 2. Spatial distribution, however, has a more similar pattern to the true state in the 1-way multivariate analysis where the patterns are smooth, while the analysis from the C-univariate data assimilation has smaller scale signals. From this figure, we can see the positive impact of wind fields on the analysis of atmospheric CO_2.

In the analysis of surface CO_2 fluxes, both analysis methods have a very good agreement with the true state. When we use a fixed inflation, there are spurious small signals spread all over even in the one-way multivariate data assimilation

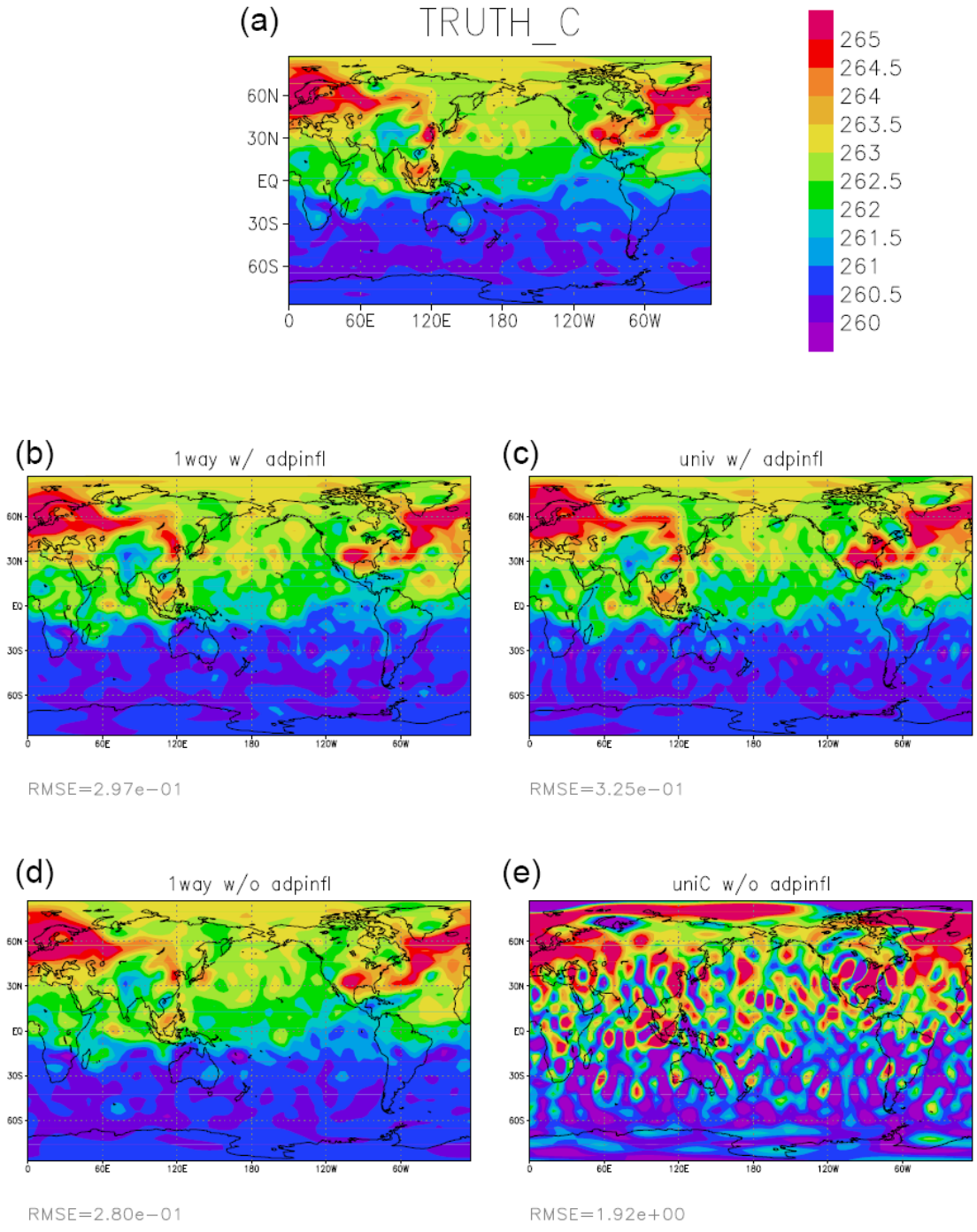

Figure 5. 2. Atmospheric CO_2 on the bottom layer: (a) True state, and analysis from (b) one-way multivariate, (c) C-univariate data assimilation with adaptive inflation, (d) one-way multivariate, (d) C-univariate data assimilation without adaptive inflation, after two months of analysis

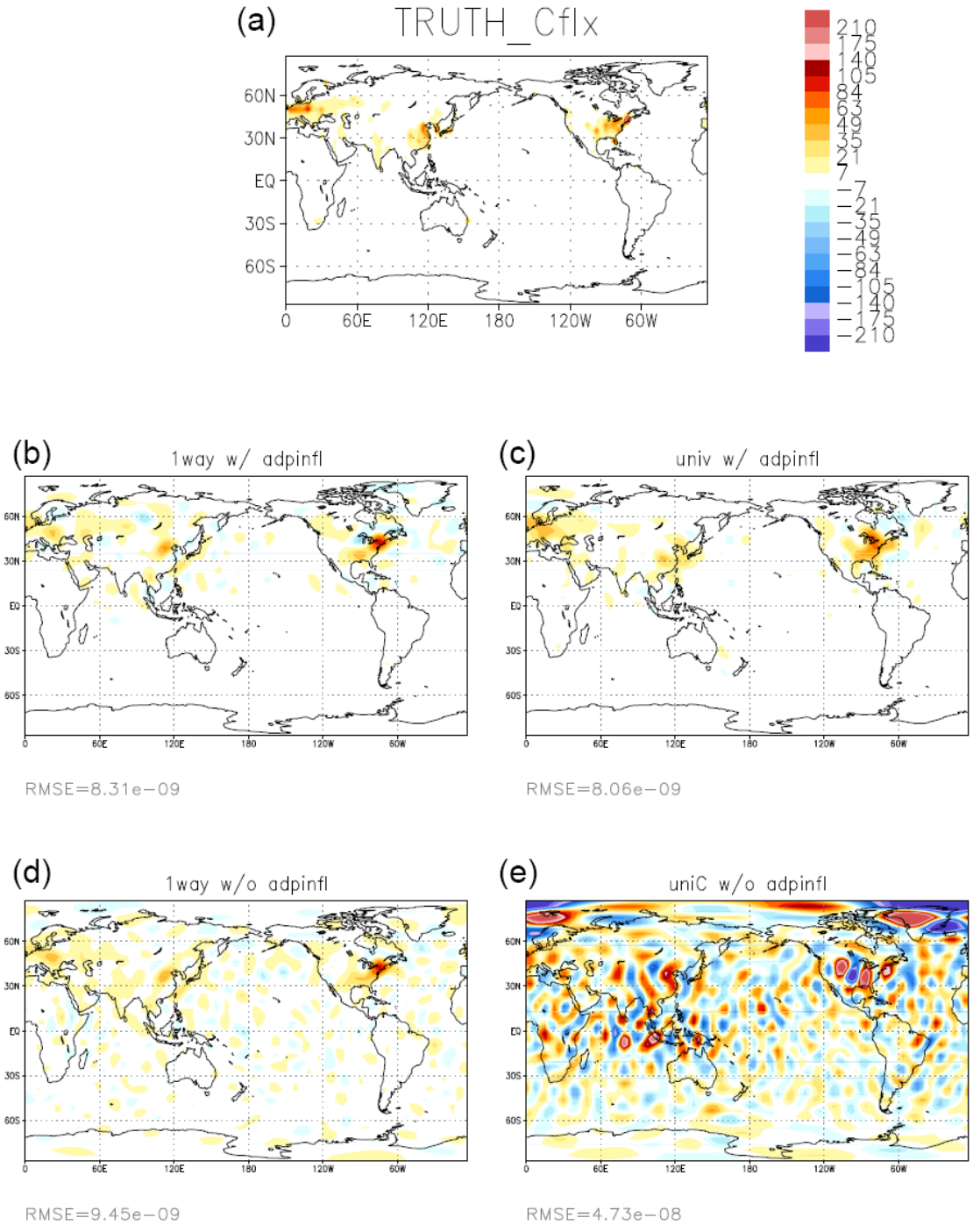

Figure 5. 3. Same as Figure 5.2, except for surface CO_2 fluxes.

(Figure 5.3 (d)). Applying the adaptive inflation, most of those small fluctuating spots disappear. On the other hand, the C-univariate data assimilation has a smaller RMS error than the 1-way multivariate. This is against our expectations based on the previous results without the adaptive inflation technique. Taking a look at the resultant analysis of Figure 5.3 (b) carefully compared to Figure 5.3 (c), we see spurious signals over northern Asia and near the major source region in US. They seem caused by the wind fields because these are prominent in the 1-way multivariate data assimilation. This suggests that the multivariate system allows for a spurious covariance between the surface carbon fluxes and the winds, and provides an idea for the next chapter on a new multivariate data assimilation system in which we perform for the first time "variable localization" in EnKF.

Next, we checked the time series of estimated adaptive inflation and observation errors (Figures 5.4-5.8). For the atmospheric CO_2 analysis, there are large values of inflation initially and then a convergence level of inflation around 5-10% for the C-univariate and 15-20% for the 1-way multivariate. Figure 5.4 also provides several of new insights. First, the C-univariate data assimilation only includes the observation of atmospheric CO_2 in the CO_2 analysis matrix so that surface CO_2 fluxes should be updated by the atmospheric CO_2 in the lowest layer. Since there is no prior information on surface CO_2 fluxes, large inflation is essential in order to give more weight to the observations than the background CO_2 that is forced by the wrong surface CO_2 flux fields at the initial time. But this large inflation is not necessary after reaching a converged state of surface CO_2 analysis. This can be

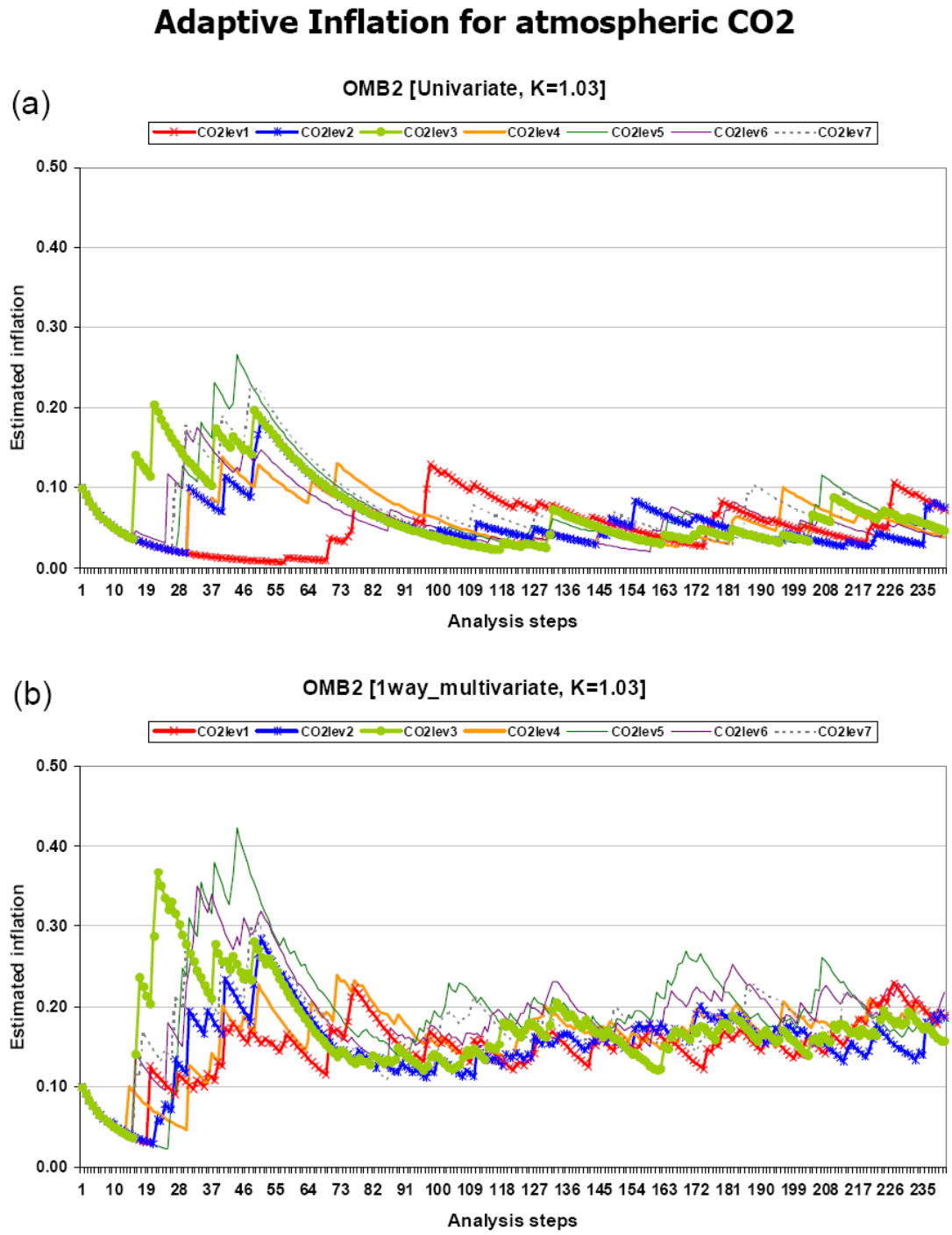

Figure 5. 4. Time series of estimated adaptive inflation for atmospheric CO_2 on each vertical layer in (a) C-univariate data assimilation and (b) one-way multivariate data assimilation through OMB^2 method. (lev1: σ =0.950, lev2: σ =0.835, lev3: σ =0.685, lev4: σ =0.510, lev5: σ =0.340, lev6: σ =0.200, lev7: σ =0.080)

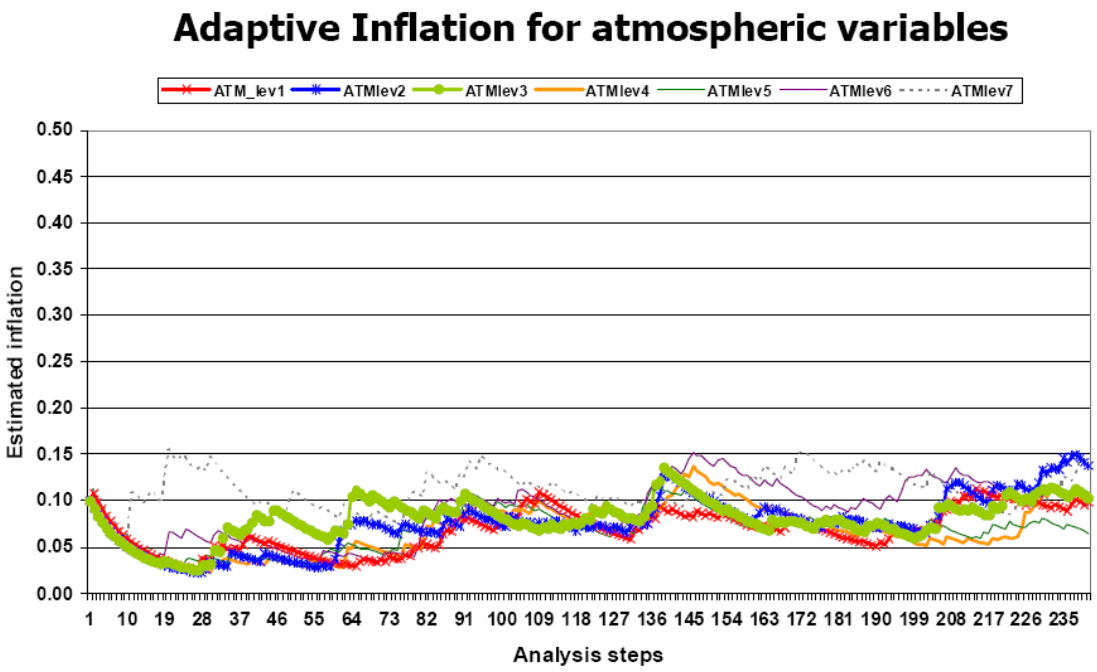

Figure 5. 5. Time series of estimated adaptive inflation for meteorological variables for each vertical layer.

supported by Figure 2.9 which shows the divergence of C-univariate data assimilation with fixed large inflation.

There are some jumps over the time series of adaptive inflation (Figure 5.4). They are related to observation error estimates. In the OSSEs, the observations for the variables are simulated with the specific standard error deviations as defined in Section 2.4.1. For example, the standard deviation of observation error in the zonal wind (U) sets 1 m/s. This means that the true observation error can be quite larger or smaller than 1 m/s over the analysis period even though it happens rarely. Then, the observation error estimates cannot catch those sudden departures and possibly affect the adaptive inflation estimate. Thus, we have some points where the adaptive inflations sometimes increase over the analysis period.

Table 5. 1. Estimated observation error standard deviations in C-univariate analysis result after two months of analysis).

Variable	True value	Estimated observation error for each vertical layer						
		level 1	level 2	level 3	level 4	level 5	level 6	level 7
U	1	1.048	1.066	1.066	1.089	1.129	1.127	1.025
V	1	1.048	1.052	1.067	1.093	1.126	1.118	1.029
T	1	1.032	1.028	1.033	1.042	1.042	1.043	1.002
q	0.1	0.119	0.12	0.112	0.111	0.109	0.103	0.098
CO2	1	1.039	1.010	1.008	1.009	1.017	1.008	1.012
Ps	100	106.020						

Table 5. 2. Same as Table 5.1, except for one-way multivariate analysis (result after two months of analysis).

Variable	True value	Estimated observation error for each vertical layer						
		level 1	level 2	level 3	level 4	level 5	level 6	level 7
U	1	1.048	1.066	1.066	1.089	1.129	1.127	1.025
V	1	1.048	1.052	1.067	1.093	1.126	1.118	1.029
T	1	1.032	1.028	1.033	1.042	1.042	1.043	1.002
q	0.1	0.119	0.12	0.112	0.111	0.109	0.103	0.098
CO2	1	0.983	0.964	0.963	0.972	0.972	0.972	0.958
Ps	100	106.020						

In addition, the reason why the resultant inflation of one-way multivariate data assimilation tends to be greater than that of C-univariate data assimilation can be explained as follows: the C-univariate data assimilation has fewer constraints than the one-way multivariate in the CO_2 analysis so that the ensemble spread does not shrink after the analysis step compared to that of the one-way multivariate system. We have already seen this feature from Figure 2.10 and 2.11 in Chapter 2. Thus, the inflation

estimated adaptively is generally larger in the 1-way than the C-univariate after the first several days. For the atmospheric variables, the estimated inflation is about 10% for both analysis systems (Figure 5.5).

Since this is a perfect model case, the estimation of observation error is more precise than that of Chapter 4 (Tables 5.1 and 5.2). All variables at all vertical levels have a very good convergence to a value close to the true value of observation error in both the C-univariate and one-way multivariate data assimilation. Figures 5.6 and 5.7 show the time series of estimated observation error, and we can see how fast the estimation converged to the true value. This fast convergence of estimated observation error was also confirmed in the imperfect model experiments.

The adaptive inflation for the surface CO_2 fluxes converges to a very small value such as 0.45% in C-univariate data assimilation and 1.13% in the one-way multivariate data assimilation (Figure 5.8). This value is far different from what we used in the experiments of Chapter 2. From this, we can say that the adaptive inflation for surface CO_2 fluxes results in a smaller magnitude than those for any other variables. Again, this is consistent with our reasoning.

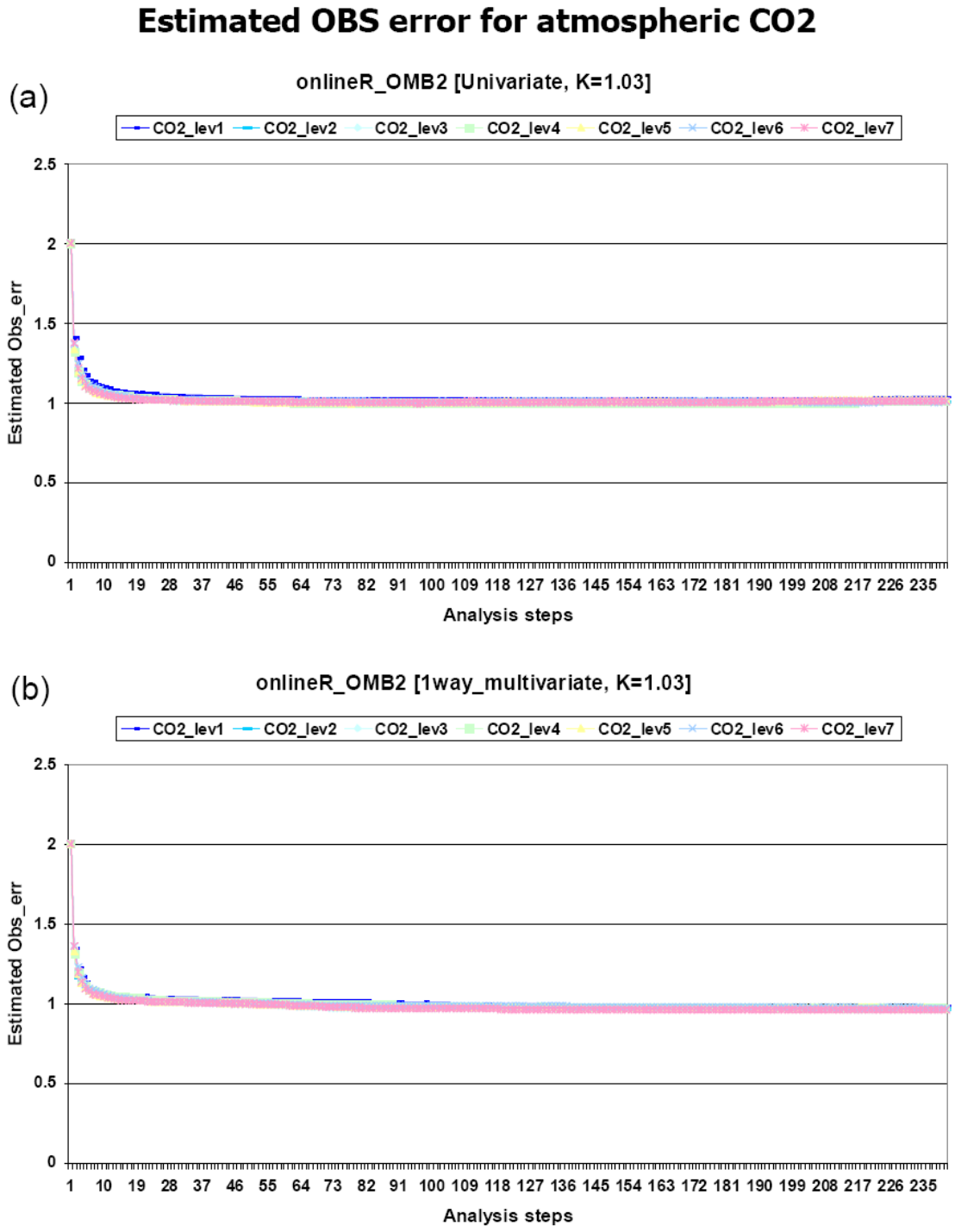

Figure 5. 6. Time series of estimated observation error for atmospheric CO_2 in (a) C-univariate, (b) one-way multivariate data assimilation for every vertical level.

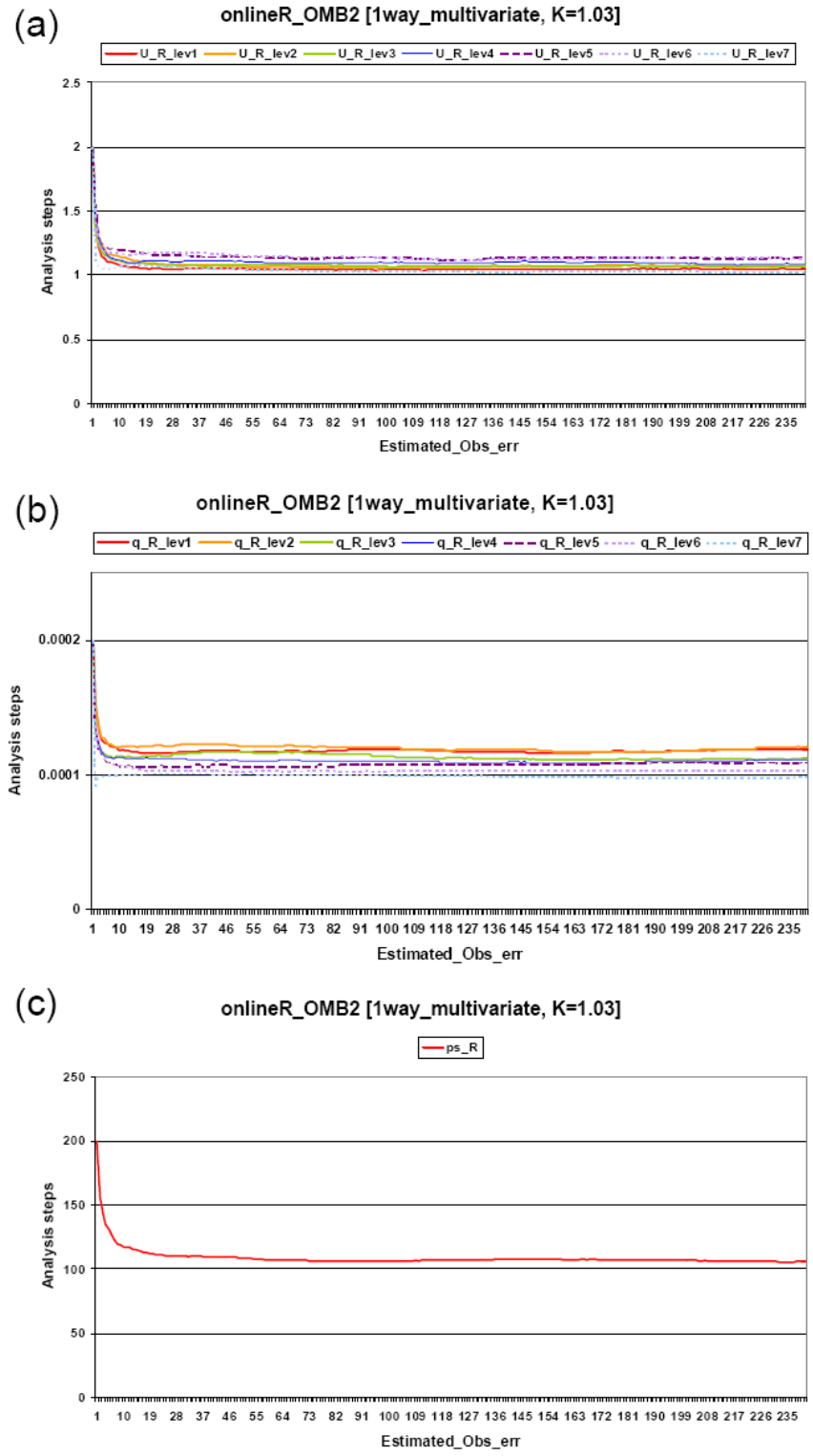

Figure 5. 7. Same as Figure 5.6, except for (a) zonal wind, (b) specific humidity, and (c) surface pressure

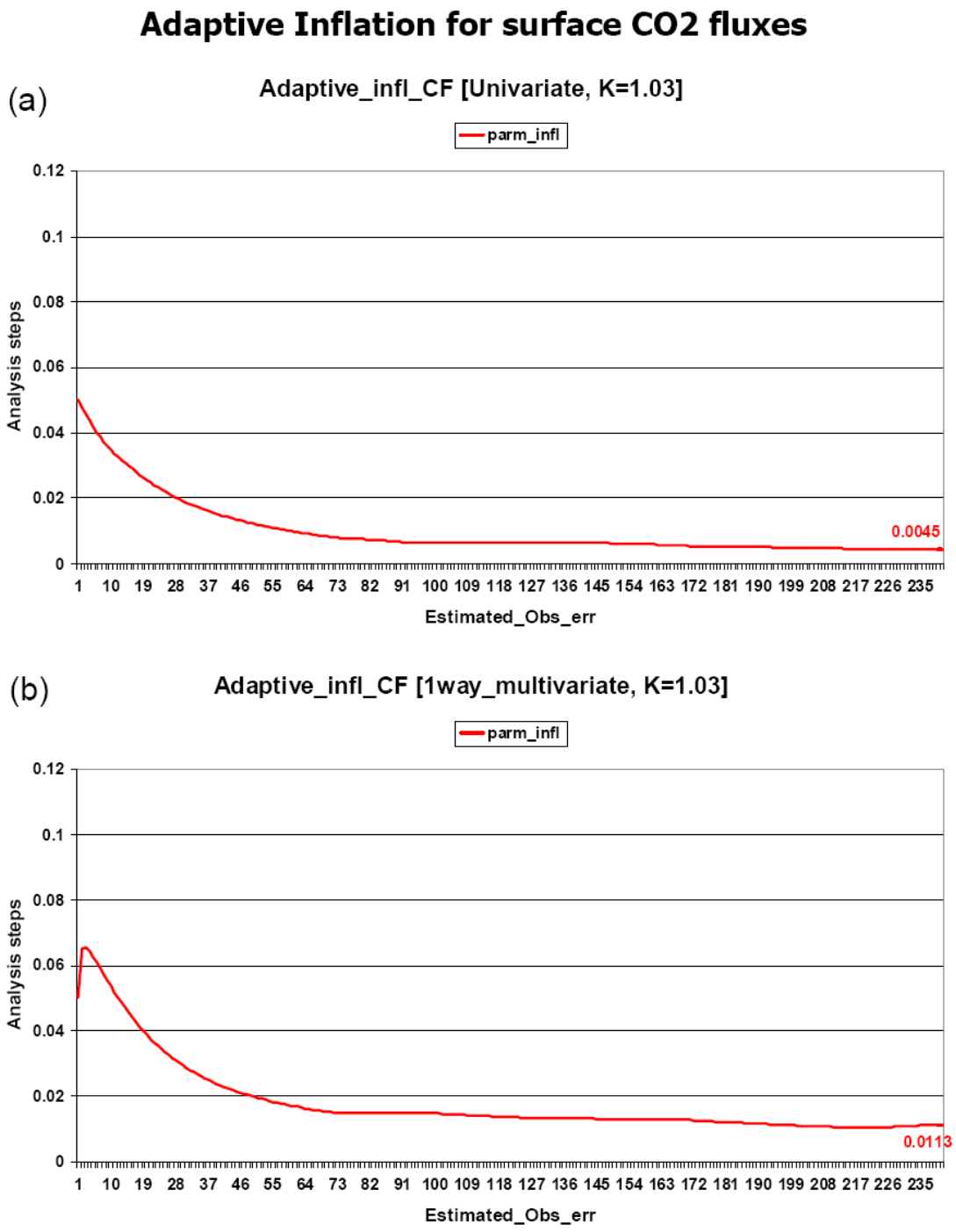

Figure 5. 8. Time series of adaptive inflation for surface CO_2 fluxes in (a) C-univariate, (b) one-way multivariate data assimilation.

5.4. Summary and discussion

We applied the adaptive inflation techniques we tried in the previous chapter to the perfect model simulation in order to see how much it could affect the C-univariate data assimilation which failed to make a two-month analysis. The results are surprisingly good for the C-univariate analysis, which are even better than the one-way multivariate data assimilation although for both techniques adaptive inflation improves the results.

Analyzing the estimated adaptive inflation, we could see that the C-univariate data assimilation requires large inflation for atmospheric CO_2 at the beginning of analysis and very small inflation for surface CO_2 fluxes in order to maintain a stable system. A time series of adaptive inflation in one-way multivariate data assimilation also shows a similar tendency, but these conditions are more needed in C-univariate data assimilation because the analysis of surface CO_2 fluxes relies on only one variable, atmospheric CO_2 in the lowest layer. Therefore, it is essential to use the observation information flexibly, especially in the C-univariate data assimilation.

From the results in this chapter, we realized that the correlation between wind fields and surface CO_2 fluxes may not be as useful as we had originally thought. Rather, spurious signals appear near the major source region since the transport of atmospheric CO_2 is significant by wind. This tells us that though the impact of wind on atmospheric CO_2 is obvious, the surface CO_2 fluxes may not be directly linked to the wind field in a physical sense.

Chapter 6: New Approach for Multivariate Data Assimilation in Carbon Cycle Data Assimilation

6.1. Introduction

In Chapter 5, we confirmed that the C-univariate data assimilation under the perfect model simulation also has a very good performance estimating surface CO_2 fluxes with a huge help from an advanced technique of adaptive inflation and observation error estimation. The surprising result that instigates this chapter is that the C-univariate data assimilation had an even better performance than the 1-way multivariate data assimilation in the analysis of surface CO_2 fluxes. This was contradictory to our expectations built on previous knowledge. What we found in the surface CO_2 analysis of 1-way multivariate data assimilation was that there were erroneous signals near the major source regions and they apparently came from their relation to the wind fields. We came to realize that surface CO_2 fluxes are not directly related to the wind fields while atmospheric CO_2 concentration on the lowest layer is obviously linked to both wind fields and surface CO_2 fluxes. That is, the 1-way multivariate data assimilation used in the previous chapters includes the covariance between surface CO_2 fluxes and wind fields. Therefore it resulted in having spurious signals over the region where the atmospheric CO_2 is dominantly transported by wind, not where CO_2 is actually released or absorbed. Thus, we developed a new multivariate data assimilation system which can include the uncertainty of wind variables to the CO_2 analysis in a physically based way.

6.2. New approach for multivariate data assimilation

It is common practice in EnKF to introduce "space localization" into the background error covariance (Houtekamer et al., 2001; Hamill et al., 2001; Gaspari and Cohn, 1999). This is because background ensemble perturbations have error covariances that are good estimates of real correlations for relatively short distances up to about 500-1000km. At longer distances, the background errors are also apparently correlated, but these correlations are just random due to sampling errors, and can seriously harm the analysis. In the widely adopted technique of "space localization" to solve the problem of long distance spurious correlations, the background error covariance terms are multiplied by a Gaussian function that decreases with the distance between the two grid points whose error covariance is being computed and becomes zero at distances longer than about 1000km.

The results that we obtained suggest applying the same concept to covariances of variables that are not physically correlated in order to reduce spurious correlations. While the atmospheric CO_2 concentration is determined by both wind fields and surface CO_2 fluxes, the surface CO_2 fluxes are not really dependent on the wind fields. Thus, we zero out the error covariance between wind fields and surface CO_2 fluxes in the previous multivariate data assimilation. This is a new methodology which we can denote "variable localization" because the system localizes the variables that the wind fields can affect. On the other hand, since C-univariate data assimilation has no correlation between CO_2 variables and atmospheric variables in the background error covariance matrix (Figure 6.1(a)), so we stick to the same formulation for comparison with the new method.

6.2.1. New 1-way multivariate data assimilation with variable localization

In the multivariate data assimilation background error covariance that we discussed in Chapter 2 and reproduce schematically in Figure 6.1, we need to modify the covariances corresponding to the one-way multivariate approach (Figure 6.1(b)) and the fully multivariate approach (Figure 6.1(c)).

In the nature, the effect of wind on atmospheric CO_2 is important but that of temperature, humidity and surface pressure is not. Similarly, the winds do not transport surface fluxes of carbon. For these reasons, we built a new analysis system based on the 1-way multivariate data assimilation: First, the system for the atmospheric variables is the same as the previous 1-way multivariate analysis (green box in Figure 6.1(b) and (d)); for updating CO_2 analysis, the background error covariance between surface CO_2 fluxes and wind field (pink box in Figure 6.1(d)) is zeroed out (pink box in Figure 6.1(d)). Thus, atmospheric CO_2 is analyzed by the error correlations among the variable itself, wind fields, and surface CO_2 fluxes, whereas the surface CO_2 fluxes are updated by the background error covariance with only atmospheric CO_2. With this system, we can still include the uncertainties of wind fields to help the analysis of atmospheric CO_2 at every vertical level. But there is no direct effect of the wind errors on the surface CO_2 fluxes.

6.2.2. New multivariate data assimilation with variable localization

In order to do “variable localization”, we have to zero out the covariance

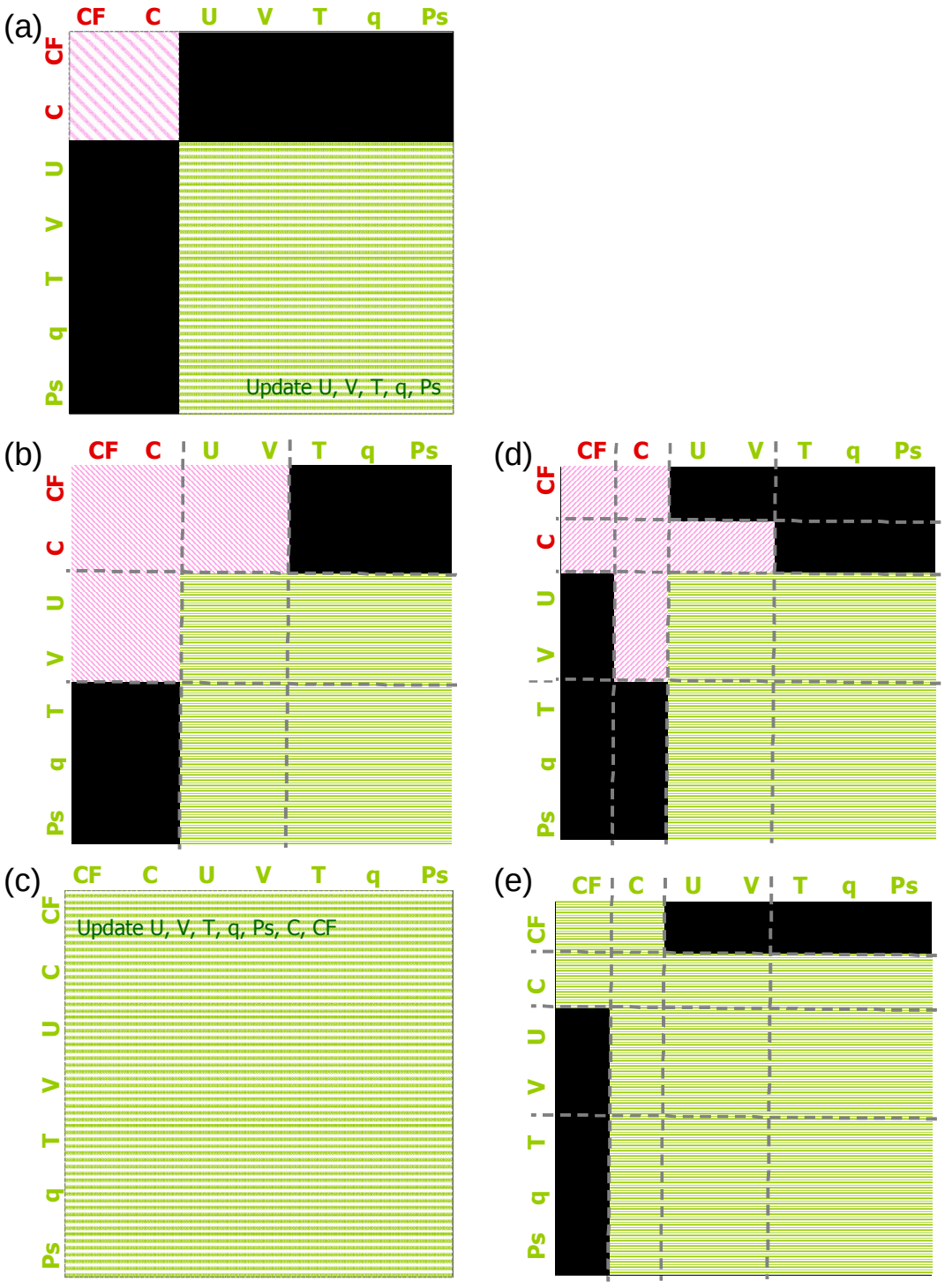

Figure 6. 1. Schematic plots of background error covariance matrix in (a) C-univariate, (b) previous one-way multivariate, (c) previous multivariate, (d) new one-way multivariate, and (e) new multivariate data assimilation.

between surface CO_2 fluxes (CF) and atmospheric variables (U, V, T, q, Ps) (Figure 6.1(e)). Thus, we let the errors of all atmospheric variables be coupled with the background error of atmospheric CO_2, but restrict the influence of atmospheric variables on surface CO_2 fluxes. For the analysis of surface CO_2 fluxes, we only include the covariance between atmospheric CO_2 on the lowest layer and surface CO_2 fluxes.

6.3. Experimental Design

We test both new multivariate data assimilations introduced above under the perfect model simulation and imperfect model simulation in that order. For the perfect model simulation, the setting is the same as Section 5.2 except for using a different inflation of atmospheric CO_2 over land and ocean as well as estimating an adaptive inflation of surface CO_2 fluxes at every grid point. Here, both adaptive inflation techniques are tested. In addition, we tried the experiment having less observation density of atmospheric CO_2 concentration; it assumes the observation every four by four grid points, so the coverage is about 6.3 % in the horizontal. On the other hand, the imperfect model simulation has been done with the same experimental design of Section 4.2.

6.4. Results

As usual, there is not significant difference between the performance of the OMB^2 and the AMB*OMB methods, therefore we show the result from only one of them.

6.4.1. Perfect model simulation with variable localization

The result from the AMB*OMB method for the adaptive inflation is shown here since it is comparable to that from the OMB^2 method. From the chart of RMS error in Figure 6.2, both new multivariate analyses have less RMS error for CO_2 variables compared to Figure 5.1. As we expected, the performance for the atmospheric variables is similar to the previous, since we did not change the error covariance matrix for these variables, but the changes in the CO_2 analysis are significant. We cannot see a visible degradation anymore in the analysis of surface CO_2 fluxes from both new multivariate data assimilation methods compared to that from the C-univariate analysis. Meanwhile, the error level is much improved in terms of CO_2 analysis compared to Figure 5.1 (e) and (f).

In case of atmospheric CO_2, the errors were about 0.3 ppmv in both C-univariate and 1-way multivariate data assimilation (Figure 5.1) and these were good enough results since the observation error is 1.0 ppmv. However, the analysis error has been further reduced to less than 0.2 ppmv for all three analysis schemes. The result that there is a significant improvement even in the C-univariate data assimilation, shows the impact of using different inflation for the CO_2 variables in the horizontal as that is the only difference between the settings of Chapters 5 and 6 in the C-univariate analysis experiments. Indeed, horizontally different inflation for CO_2 also helped the CO_2 analysis under the imperfect model simulation in Chapter 4, because the variability of CO_2 over land is much larger than that over the ocean as we discussed.

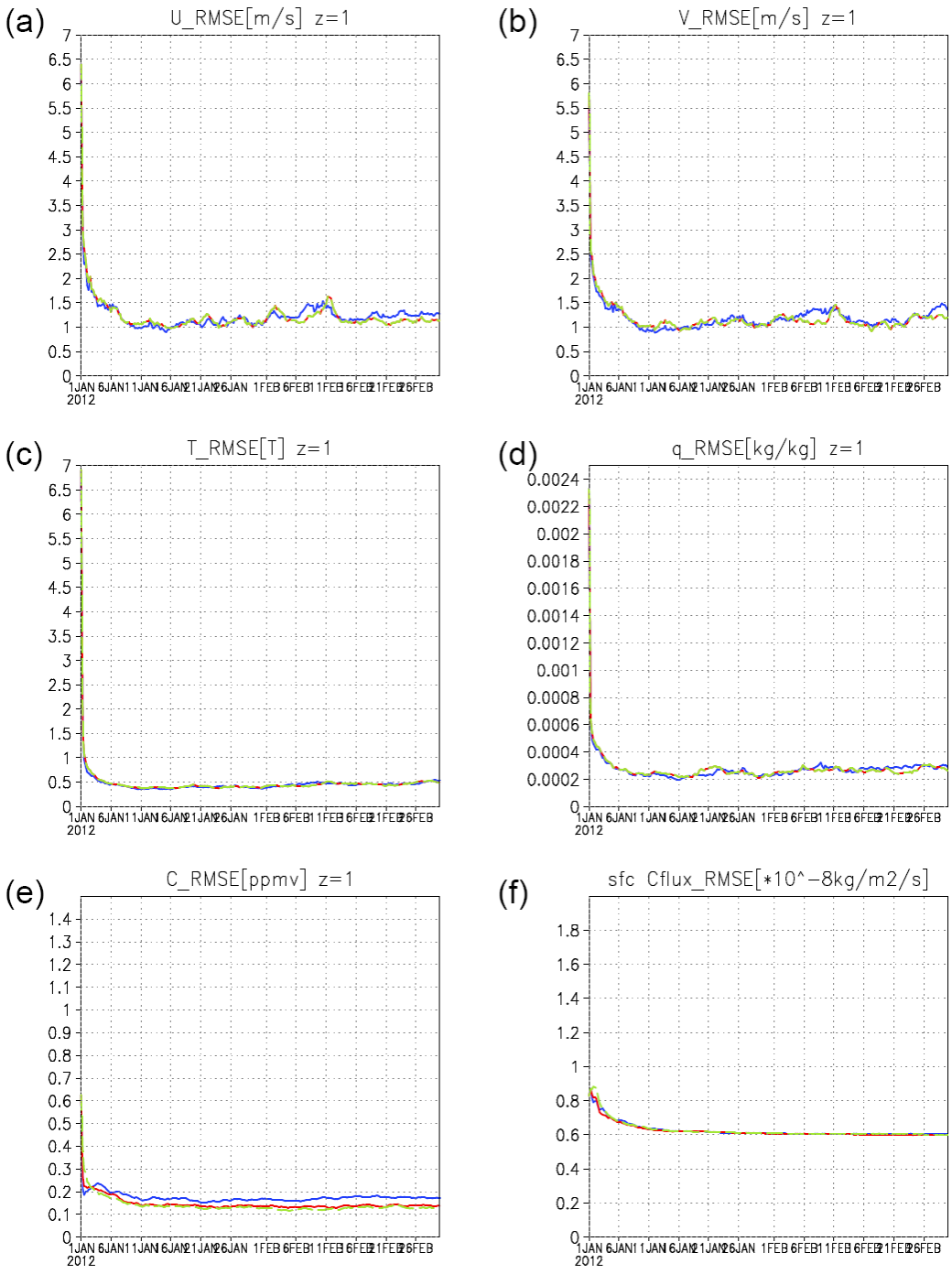

Figure 6. 2. RMS errors of (a) U (m/s), (b) V (m/s), (c) T (K), (d) q (kg/kg), (e) atmospheric CO_2 on the lowest layer, (f) surface CO_2 fluxes. Green indicates C-univariate data assimilation, red results from new one-way multivariate analysis, and blue from new multivariate analysis.

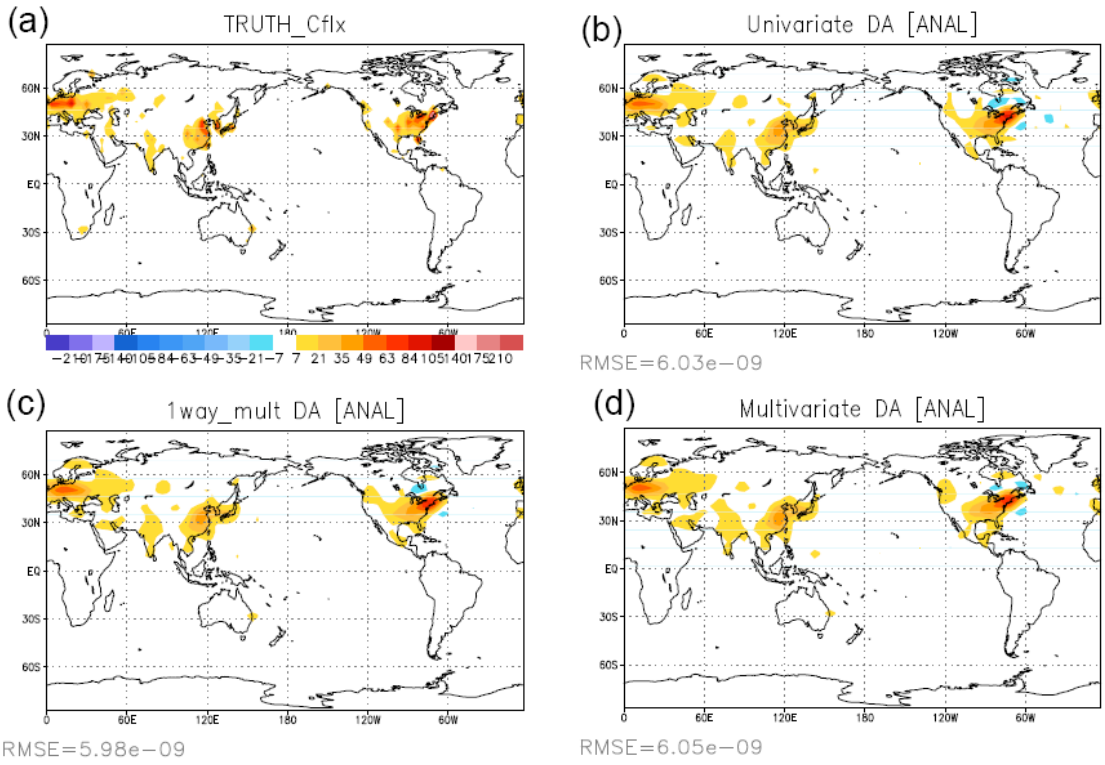

Figure 6. 3. surface CO_2 fluxes: (a) True state, analysis from (b) the C-univariate, (c) the new one-way multivariate, and (d) the new multivariate data assimilation after two months of analysis. (unit: 10^{-9} kg/m^2/s)

The differences of Figure 6.2(e) from Figure 5.1(e) verify the importance of horizontally different inflation for CO_2 data assimilation.

Returning to the performance of the new multivariate data assimilation, Figure 6.3 shows that the new 1-way multivariate data assimilation has less RMS error at the end of the analysis and the spatial distribution is also closer to the true state than with the C-univariate. For example, the new 1-way multivariate data assimilation does have less spurious forcing near the eastern US. This can be attributed to the uncertainty of wind over these regions helping correct the atmospheric CO_2 fields, so that 1-way multivariate data assimilation has now a signal of surface CO_2 flux analysis closer to the true state. That is, since the 1-way multivariate system contains

the correlation between atmospheric CO_2 and wind, correction of the atmospheric CO_2 analysis owing to the wind error information can be emphasized around major source regions and where the network of wind observation is dense enough. Moreover, only both multivariate data assimilation are able to catch the very small source over eastern Australia. These results, we believe, convincingly shows that the new 1-way multivariate system developed here is the best algorithm we can find to estimate surface CO_2 fluxes.

Reducing the observation density from 25% to 6.3%, the observations are 15° far from each other in longitude. The analyses of surface CO_2 from both the C-univariate and the one-way multivariate data assimilation results in slightly larger RMS errors (Figure 6.4) than the case with dense observation network. Still, the one-way multivariate analysis has a smaller error than the C-univariate although the difference is not significant. Besides, the signals start to become somewhat noisy near major source regions. However, the error levels and the spatial distributions are still comparable to the true states.

6.4.2. Imperfect model simulation with variable localization

In general, the results from the OMB^2 method is usually similar to that from the AMB*OMB method in terms of RMS errors in the imperfect model experiment. We will show results from the experiment with the OMB^2 method. There is a comparison of CO_2 analyses among the previous 1-way multivariate analysis, the new 1-way multivariate, the previous multivariate, the new multivariate, and the

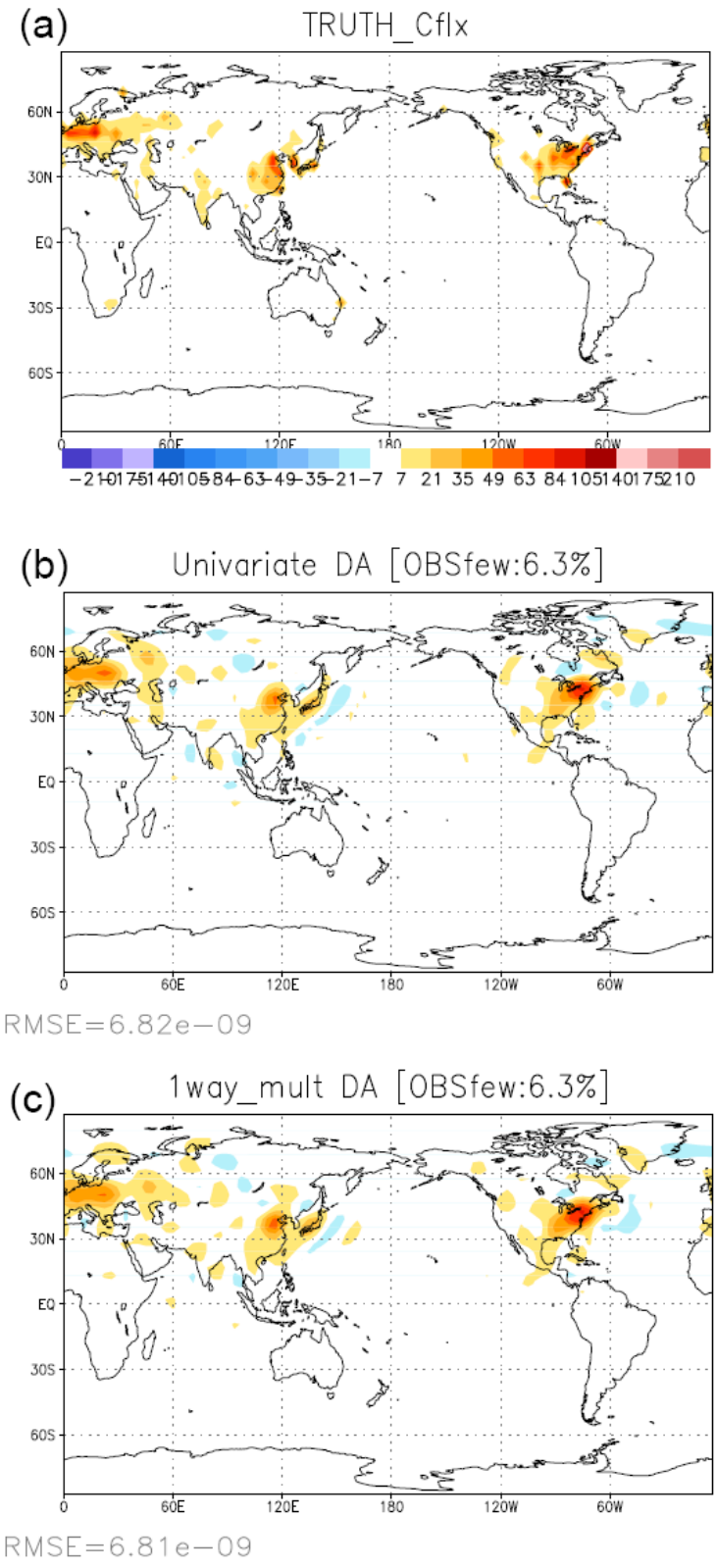

Figure 6. 4. (a) True state of surface CO_2 fluxes after two months of analysis, and the analysis of surface CO_2 fluxes in (b) C-univariate , and (c) one-way multivariate data assimilation with the CO_2 observation at every four grid point (6.3% coverage).

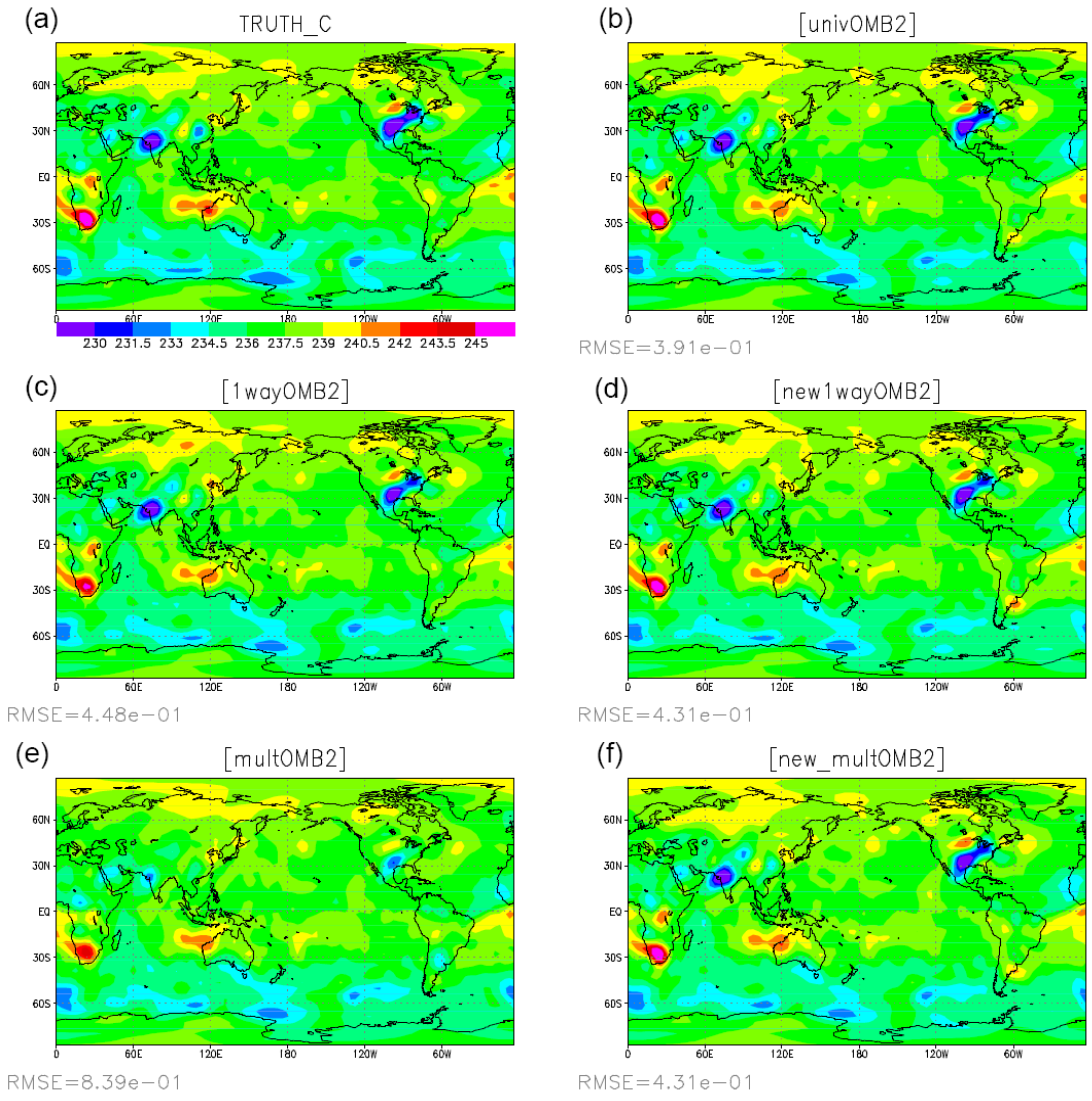

Figure 6. 5. Atmospheric CO_2 on the bottom layer: (a) truth, analysis in (b) C-univariate, (c) old one-way multivariate, (d) new one-way multivariate, (e) old multivariate, (f) new multivariate data assimilation.

C-univariate analysis (Figures 6.5 and 6.6) in terms of both RMS error and spatial distribution. We confirmed for the imperfect model that both of the new multivariate data assimilation systems also estimate better the spatial distribution for CO_2 variables than the previous multivariate systems.

First, one can see that the atmospheric CO_2 analysis is improved after implementing the new 1-way multivariate data assimilation when comparing Figure 6.5(c) to Figure 6.5(d). Again, since there are relatively dense observations of atmospheric CO_2, the analysis of CO_2 from all the methods (Figure 6.5 (b), (d), and (f)) are good enough and the difference among them is not significant: 0.04 ppmv. However, the improvement on the fully multivariate assimilation from Figure 6.5 (e) to Figure 6.5 (f) is remarkable and this tells us that negative impact on the CO_2 analysis can be caused by coupling the errors of irrelevant variables such as temperature, humidity, surface pressure.

From the analysis of surface CO_2 fluxes, one can clearly see the impact of variable localization introduced in the new multivariate system. Taking a look at Figure 6.6 (d), and (f), the contamination from the coupled error of surface CO_2 fluxes with other atmospheric variables is reduced significantly in both multivariate data assimilation systems without variable localization (compare to Figure 6.6 (c) and (e)). It is evident that the deviation of analysis from the true state is more serious in the multivariate data assimilation than the 1-way multivariate one, especially over Africa, northern Europe, and northern Asia. These features convince us that the new concept of multivariate data assimilation with variable localization introduced here is

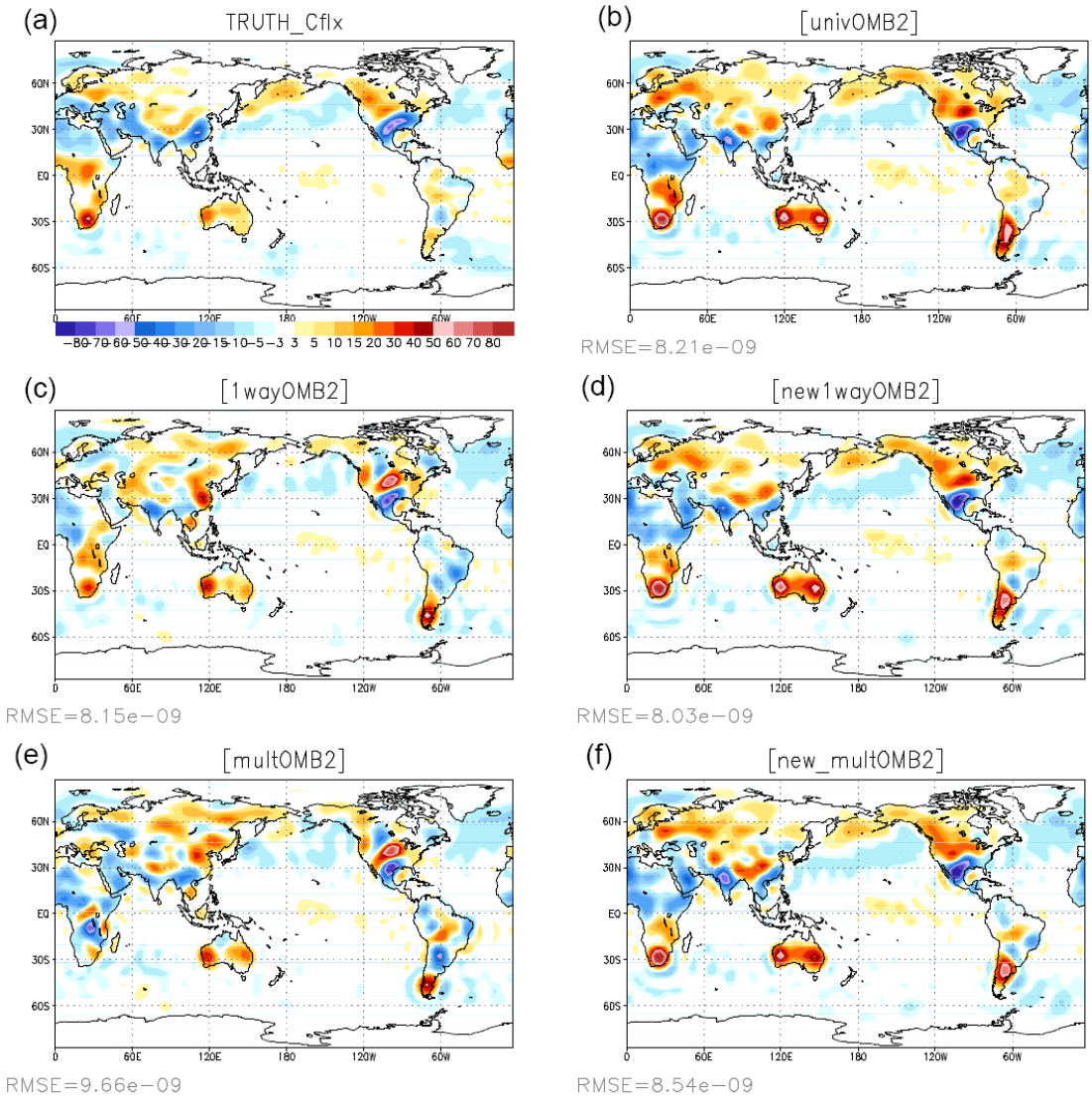

Figure 6. 6. Same as Figure 6.5, except for surface CO_2 fluxes.

an important finding in estimating surface CO_2 fluxes by reflecting information of atmospheric variables in a physically meaningful way.

Indeed, the improvement of CO_2 flux analysis is very significant. With the new idea of multivariate analysis with variable localization, 1-way multivariate data assimilation has an optimal performance on estimating surface CO_2 fluxes in RMS error and spatial distribution. Moreover, the multivariate analysis of Figure 6.5(f) is much better than the previous multivariate analysis which allows a correlation with errors of all variables. That is, unphysical correlation among irrelevant variables caused a degradation of analysis and this was more serious in the previous multivariate analysis rather than 1-way multivariate data assimilation because the one-way multivariate reduced the sampling error in the covariances. This result is clearly consistent with our understanding.

The simultaneous estimation of the true value of observation errors in the new multivariate systems is as good as before for all of three methods (Table 6.1, 6.2 and 6.3). Since the atmospheric analysis of 1-way multivariate data assimilation should be the same as the C-univariate one, we confirmed that the estimation of observation error and adaptive inflation gives the same results for the atmospheric variables. From the charts of adaptive inflation for atmospheric CO_2 (Figure 6.7, 6.8, and 6.9), we could see that the multivariate data assimilation tends to have larger inflation than the 1-way multivariate, which has also slightly larger inflation than the C-univariate one. This is because the multivariate data assimilation has more constraint in the analysis of CO_2 so that the system requires larger inflation to reduce the sampling error.

Table 6. 1. Estimated observation error standard deviations in new one-way multivariate analysis

Vari-able	True value	Initial Guess	Estimated observation error for each vertical layer						
			level 1	level 2	level 3	level 4	level 5	level 6	level 7
U	1	2	1.076	1.111	1.091	1.168	1.221	1.208	1.026
V	1	2	1.042	1.108	1.117	1.159	1.220	1.192	1.055
T	1	2	1.059	1.017	1.059	1.056	1.075	1.068	1.008
q	0.1	0.2	0.272	0.165	0.120	0.114	0.112	0.102	0.100
CO2	1	2	1.000	0.959	0.962	0.973	0.964	0.974	0.972
Ps	100	200	108.043						

Table 6. 2. Same as Table 6.1, except for C-univariate data assimilation

Vari-able	True value	Initial Guess	Estimated observation error for each vertical layer						
			level 1	level 2	level 3	level 4	level 5	level 6	level 7
U	1	2	1.076	1.111	1.091	1.168	1.221	1.208	1.026
V	1	2	1.042	1.108	1.117	1.159	1.220	1.192	1.055
T	1	2	1.059	1.017	1.059	1.056	1.075	1.068	1.008
q	0.1	0.2	0.272	0.165	0.120	0.114	0.112	0.102	0.100
CO2	1	2	1.000	0.946	0.949	0.954	0.946	0.957	0.960
Ps	100	200	108.043						

Table 6. 3. Same as Table 6.1, except for multivariate data assimilation

Vari-able	True value	Initial Guess	Estimated observation error for each vertical layer						
			level 1	level 2	level 3	level 4	level 5	level 6	level 7
U	1	2	1.118	1.134	1.131	1.204	1.268	1.249	1.045
V	1	2	1.064	1.139	1.134	1.206	1.272	1.216	1.085
T	1	2	1.074	1.035	1.049	1.079	1.082	1.072	1.031
q	0.1	0.2	0.295	0.173	0.126	0.125	0.117	0.104	0.102
CO2	1	2	1.000	0.979	0.975	0.981	0.978	0.986	0.980
Ps	100	200	108.968						

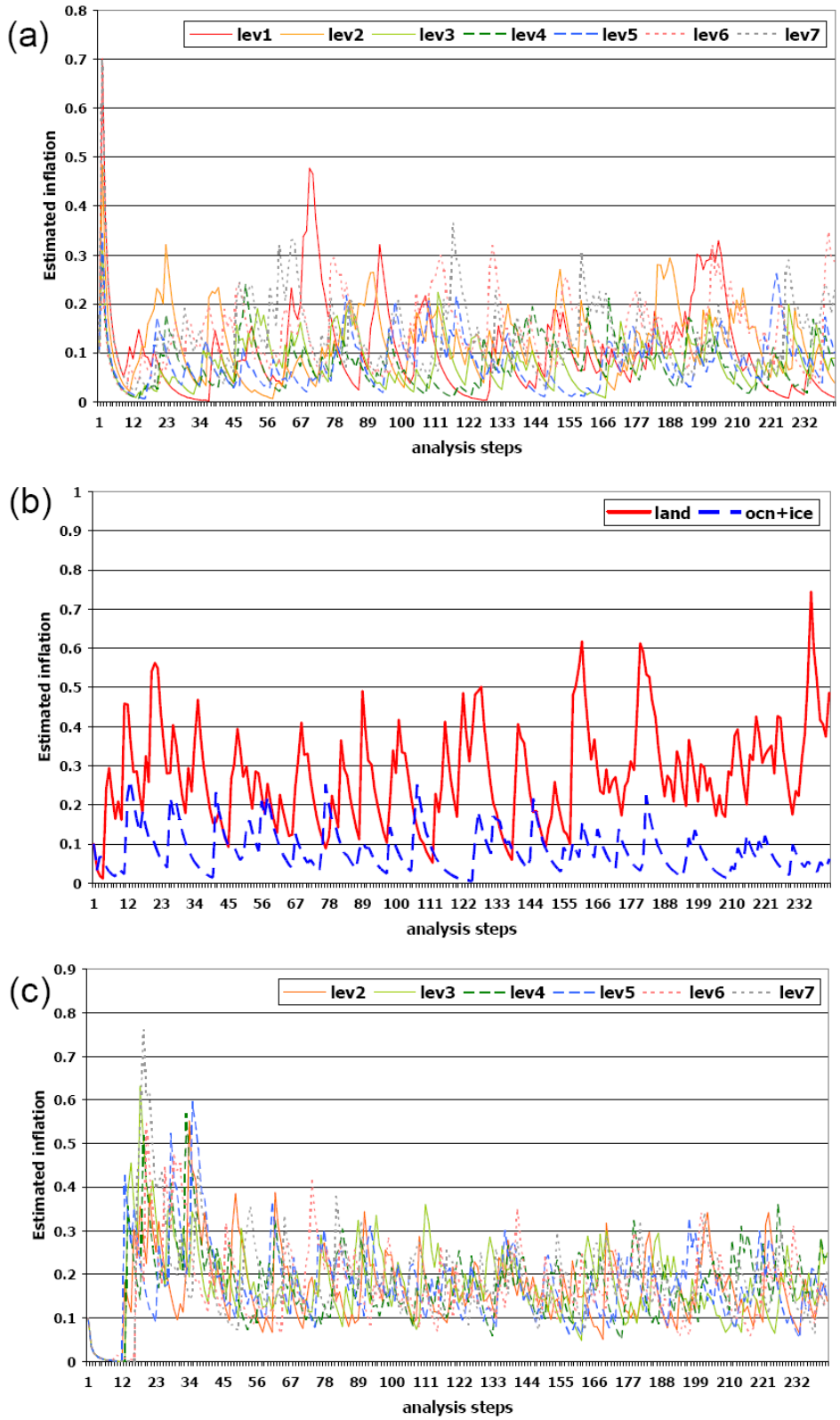

Figure 6. 7. Time series of resultant adaptive inflations ($\Delta - 1$) for (a) meteorological variables for all vertical levels, (b) atmospheric CO_2 on the bottom layer, and (c) atmospheric CO_2 on upper levels, in the C-univariate data assimilation

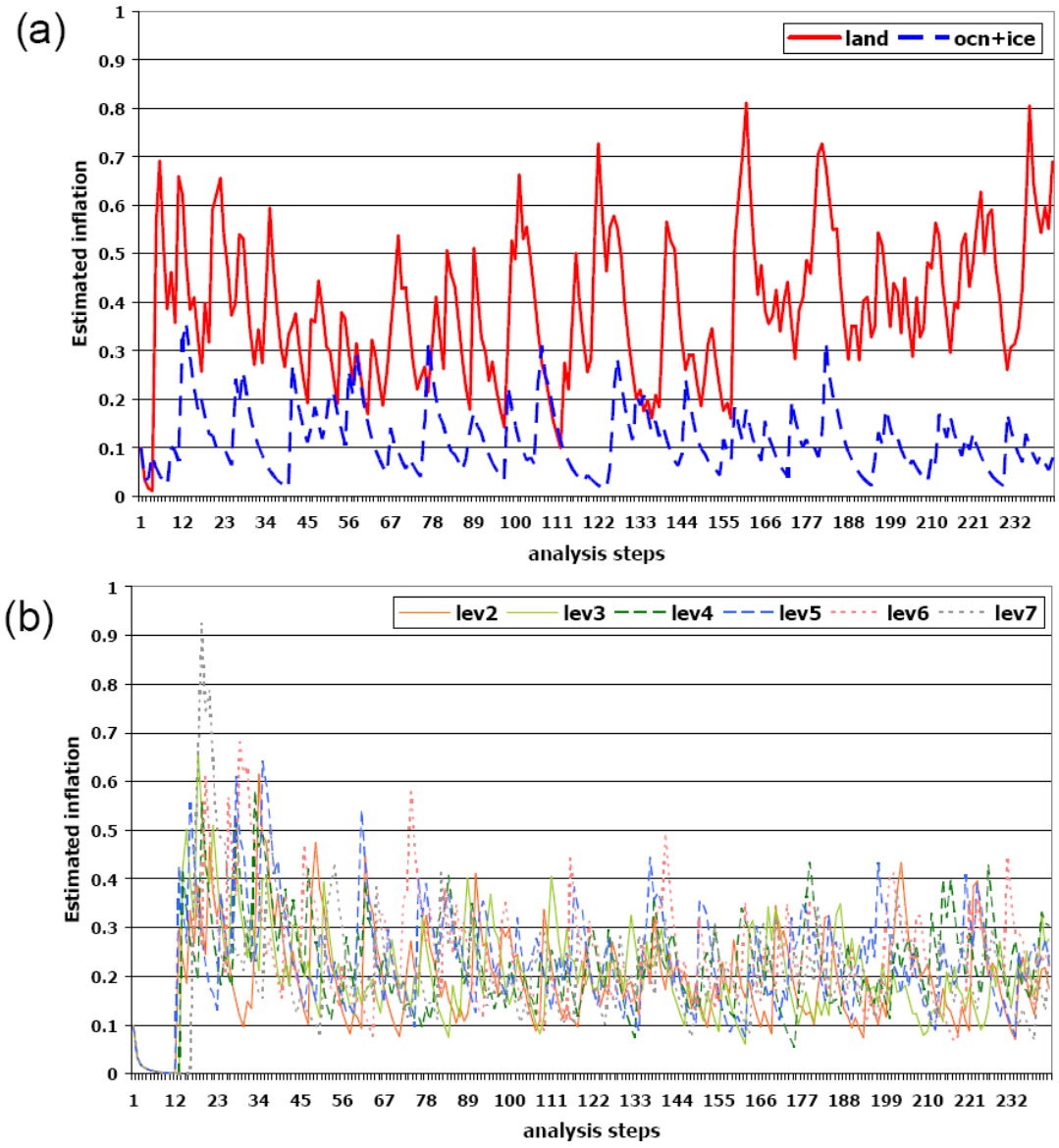

Figure 6. 8. Time series of resultant adaptive inflations ($\Delta - 1$) for (a) atmospheric CO_2 on the bottom layer, and (b) atmospheric CO_2 on upper levels, in the new one-way multivariate analysis.

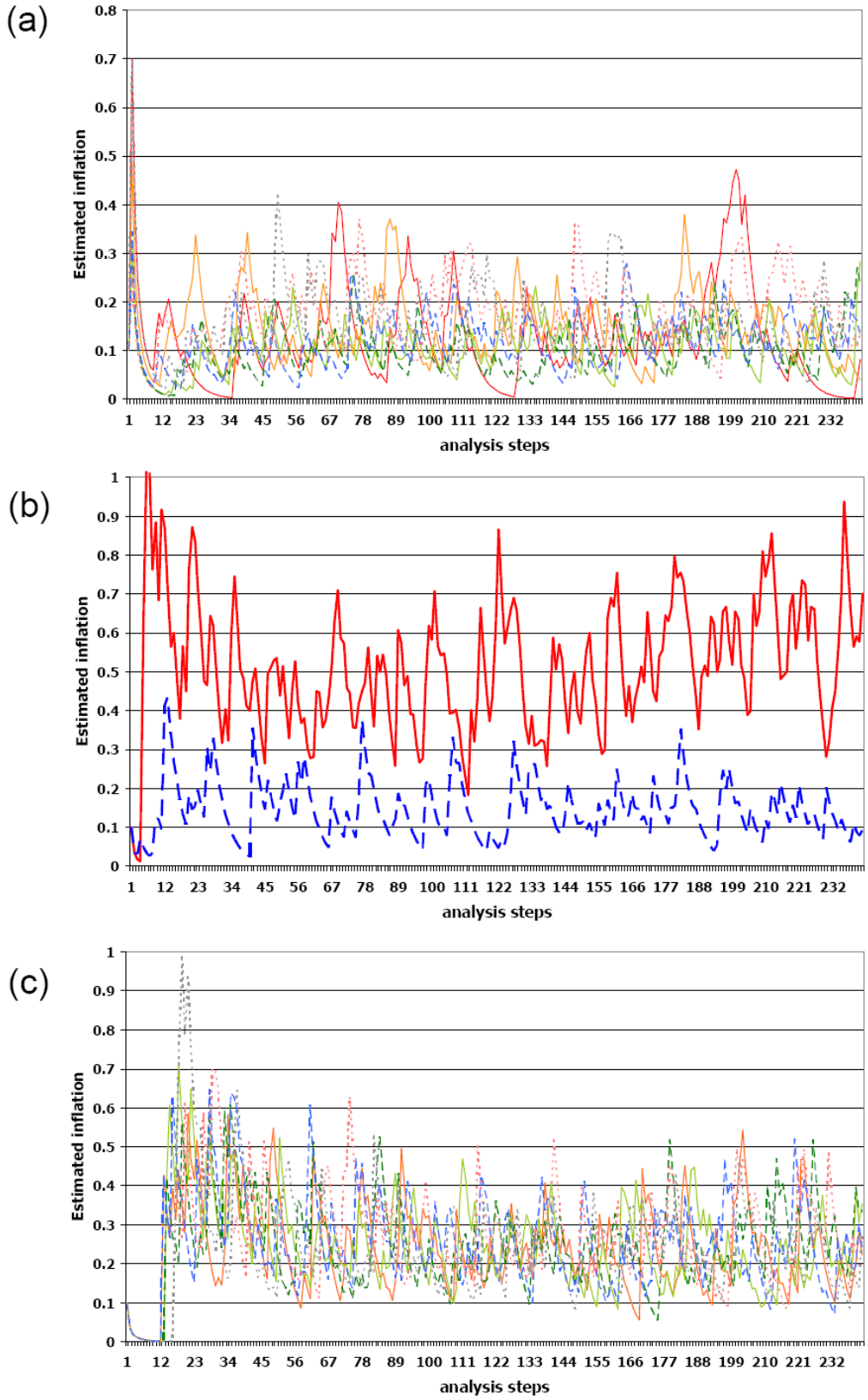

Figure 6. 9. Same as Figure 6.7, except for the new multivariate data assimilation.

6.5. Summary and discussion

From the work of the previous chapter, we found that the analysis of surface CO_2 fluxes in one-way multivariate analysis was slightly worse than that in C-univariate analysis, a result that we tracked down to the coupling in the background error covariance of surface carbon fluxes and wind fields. Thus, we developed a new concept of both one-way multivariate and multivariate data assimilation which removes the correlation between surface CO_2 fluxes and wind fields ("variable localization"). Then, we tested these new analysis methods in both a perfect model case and an imperfect model simulation.

In the perfect model simulation, we could see an obvious improvement of both new multivariate and new one-way multivariate data assimilations from the old multivariate ones when we did not allow the background error of meteorological variables to be coupled with surface CO_2 fluxes. However, the wind field still provides its error information to analyze atmospheric CO_2 concentration in both multivariate analysis systems. Therefore, the analysis of surface CO_2 fluxes in the new one-way multivariate data assimilation becomes better than that of the C-univariate, although the results from the C-univariate were already good. This is because the background error of the wind field helps analyze atmospheric CO_2 where the atmospheric CO_2 is transported near the major source regions. That is why a new one-way multivariate has less incorrect signals, especially near the eastern US where the old one-way multivariate has more spurious signals.

Under the imperfect model simulation, both new multivariate data assimilation methods also work well and clearly improve the spatial distribution of

surface CO_2 flux analysis. While the new one-way multivariate data assimilation has the best results, the improvement on the multivariate data assimilation with the new concept is remarkable. This result is consistent with our knowledge because the previous multivariate analysis includes a large portion of interacting errors among wind, temperature, humidity, surface pressure, atmospheric CO_2 and surface CO_2 fluxes. The reason why the multivariate data assimilation is still worse than the one-way multivariate data assimilation is that error of atmospheric CO_2 is still coupled with temperature, humidity and surface pressure so that these correlations do not help analyze surface CO_2 fluxes.

We obtained a significant improvement of 1-way multivariate data assimilation in the experiments allowing for no correlation between wind errors and surface CO_2 flux errors, which is a valid assumption under our experimental setting. In these experiments, there is no coupling between atmosphere and ocean and we use a prescribed oceanic CO_2 flux. In reality, however, the error correlation between the wind fields and the surface CO_2 fluxes over the ocean can be important because the air-sea net flux of CO_2 strongly depends on the surface wind speed (Takahashi et al., 2002; Feely et al., 2004). Thus, we may need to consider the error correlation between wind fields and surface CO_2 fluxes over the ocean for a realistic case. Still, the variable localization tried in this chapter is very valuable finding, which is useful for the case where we need to neglect the error covariance between irrelevant variables and thus reduce substantially sampling errors.

Lastly, there is one problem remaining: the appearance of “hot spots” in the surface fluxes of carbon over land in the Southern Hemisphere (Figure 6.6), which

was a feature observed mostly in the C-univariate data assimilation, before we implemented the new multivariate systems. With the new multivariate analysis methods, these hot spots also appear in the results from both multivariate data assimilations. This is an interesting case of “local filter divergence” where the estimation of the surface fluxes keeps slowly growing in time. We have tracked this problem to model bias in carbon, which we have not corrected, and will discuss how we believe it can be handled in the future plans of Chapter 7.

Chapter 7: Summary and Lessons Learned

The purpose of this study is to evaluate the feasibility of estimating surface CO_2 fluxes by assimilating remotely sensed atmospheric CO_2 observations without *a-priori* information. For the Observing System Simulation Experiments, we developed two forecast systems: SPEEDY-C and SPEEDY-VEGAS. With these systems, we developed or applied many advanced techniques that we found necessary in order to produce accurate analysis of surface CO_2 fluxes. We obtained encouraging results using the Local Ensemble Transform Kalman Filter (LETKF, Hunt et al., 2007) combined with these additional techniques. This work is part of a collaborative project with Prof. Inez Fung and Dr. Junjie Liu, who are developing a "real" system coupling the LETKF with the CAM model and will use our results to assimilate real observations from AIRS and GOSAT.

7.1. Development of SPEEDY-C and SPEEDY-VEGAS

In order to simulate atmospheric CO_2 concentration, we first developed the SPEEDY-C model which is modified from SPEEDY (Molteni, 2003). We added a tracer for atmospheric CO_2 with only two processes: advection and diffusion. SPEEDY-C does not have any physical process for surface CO_2 fluxes, so we assume it is constant (use a persistence forecast for the surface CO_2 fluxes). Thus, SPEEDY-C reads the surface CO_2 fluxes as a forcing term and the released CO_2 is transported and mixed in the atmosphere. Next, SPEEDY-C was coupled with a terrestrial

carbon model (VEGAS: Zeng et al., 2005) and physical land model (SLand: Zeng et al., 2000). With this coupled SPEEDY-VEGAS model, we could compute the time-varying CO_2 fluxes over land based on the interaction with atmospheric conditions, as well as do experiments of data assimilation under the imperfect model assumption.

7.2. C-univariate vs. multivariate data assimilation with a fixed inflation factor

Within the LETKF framework, it is possible to couple the background (forecast) errors of variables which helps the analysis of variables related to each other in nature. Thus, we developed several types of multivariate analysis systems for the CO_2 data assimilation and compared them to the C-univariate data assimilation. In C-univariate data assimilation, the analysis system for the atmospheric CO_2 and surface CO_2 fluxes assimilates only the observations of atmospheric CO_2. Thus, the Kalman gain for the atmospheric CO_2 and surface CO_2 fluxes are determined by only the atmospheric CO_2 concentration on the lowest layer. We note that until now, only C-univariate approaches have been used in both simulation and real observation data assimilation or inversion studies of atmospheric CO_2 and surface carbon fluxes.

Multivariate data assimilation allows the background errors of variables to be coupled in the analysis. In the "one-way multivariate" data assimilation, the errors of wind fields are provided in the analysis for the atmospheric CO_2 and surface CO_2 fluxes, whereas the analysis of winds and other meteorological variables are not affected by CO_2 variables. This reduces the impact of sampling errors in CO_2 on the atmospheric variables. By contrast, the multivariate data assimilation allows the

interaction of errors from all the variables in the analysis, so the analysis can include unnecessary or unphysical correlations among the variables in a physical sense, for example, the relationship between the surface CO_2 fluxes and surface pressure.

Results show that the one-way multivariate data assimilation has the best overall performance in analyzing CO_2 variables. That is because it makes full use of beneficial information given to the CO_2 analysis which is the errors of wind fields, in addition to those of atmospheric CO_2 concentration. When we use a fixed inflation factor, one-way multivariate analysis performs better than the C-univariate data assimilation, which blows up. An advantage of one-way multivariate data assimilation is to constrain the CO_2 analysis with the information of wind field errors in addition to the atmospheric CO_2 fields. For a similar reason, the multivariate data assimilation works better than the C-univariate data assimilation. That is, the multivariate data assimilation has more constraints to CO_2 variables from observations of all the variables because the errors are coupled in this analysis method. Thus, it helps the analysis to not blow up although the coupled errors do not have a meaningful relationship. Because one-way multivariate data assimilation prevents some irrelevant correlation among the variables, the multivariate data assimilation has a worse analysis of CO_2 variables than the one-way multivariate analysis. In the nature run, atmospheric CO_2 is determined by wind fields through the transport and mixing, but does not change any other atmospheric variables. The methodology of the one-way multivariate data assimilation fits the physics in the model.

7.3. Bias correction, adaptive inflation and observation error estimation

For the imperfect model simulation, we use the coupled atmosphere-vegetation-land model for the nature run and the SPEEDY-C for the ensemble forecasts. Since the climatology of the forecast model is significantly different from that of the nature run, we applied a bias correction to the standard atmospheric variables, and adaptive inflation to the LETKF for the carbon data assimilation. With the bias correction, the analyses of atmospheric variables and atmospheric CO_2 concentration improved significantly, but the surface CO_2 fluxes were not estimated properly. Implementing an adaptive inflation technique introduced by Li et al. (2009), the analysis estimates the surface CO_2 fluxes very well, while the analysis of atmospheric variables and atmospheric CO_2 were further improved. In addition, the method of adaptive inflation used for this study estimates the observation error simultaneously, and the accuracy of estimated observation error is satisfactory. This is important for CO_2 applications since it is not clear what the accuracy of remotely sensed atmospheric CO_2 measurements will be.

From the results using bias correction and adaptive inflation, we learned many things: (i) The analysis of atmospheric CO_2 requires relatively large inflation in the initial stage because the random initial condition of surface CO_2 fluxes can make the ensemble forecast far from the true state. Thus, large inflations let the analysis reflect the observation more than the ensemble forecast until the analysis makes a reasonable estimate of surface CO_2 fluxes so that the ensemble forecast can produce reasonable fields of atmospheric CO_2. (ii) It is better to estimate independent inflation factors for the atmospheric CO_2 over the land and over the ocean at the lowest layer. Since the

dominant variation of surface CO_2 fluxes occurs over the land, it is necessary to take into account the observations of atmospheric CO_2 over land more than over the ocean. However, the analysis of wind tends to reduce the ensemble spread of atmospheric CO_2 over the land more than over the ocean, because the observations of the wind field are located mainly over the land. Thus, the separate adaptive inflation of atmospheric CO_2 over the land and over the ocean has a better performance on estimating surface CO_2 fluxes. The estimated inflation results have larger values over the land than over the ocean as we expected. (iii) The inflation for the surface CO_2 fluxes should be less than the other variables. We experienced that the analysis has blown up when we used any similar or larger inflation factors for the surface CO_2 fluxes compared to those for other variables. Mathematically, we have proven that the inflation of variables having no observations should be smaller than those having observations. Then, a simple adaptive inflation method for the surface CO_2 fluxes has been examined and the results confirm that the inflation is smaller than that of other variables.

7.4. Variable localization in the multivariate data assimilation

When the adaptive inflation technique was applied to the perfect model experiment, we found the contamination caused by the direct correlation between wind fields and surface CO_2 fluxes. This contamination makes the performance of the one-way multivariate data assimilation worse than the C-univariate data assimilation. Thus, we developed a new multivariate analysis approach which we denote “variable localization” where the correlations between the errors in

atmospheric variables and in surface CO_2 fluxes (which are physically implausible) are zeroed out. In the new one-way multivariate data assimilation with the variable localization, the error information for the wind fields is used to analyze atmospheric CO_2, while only the error information of the atmospheric CO_2 concentration is used to update the surface CO_2 flux analysis. Likewise, new multivariate data assimilations with variable localization were tested, blocking the correlation of errors in the surface CO_2 fluxes with errors in all the atmospheric variables in the analysis. Still, the new multivariate analysis includes some unnecessary correlations among variables: for example, coupled errors between the atmospheric CO_2 and humidity.

Table 7.1 summarizes the methods we have examined. If one reads the row starting from W, it can be explained as follows: the wind fields (W) are analyzed using the errors of the wind fields (W) coupled with those of temperature, specific humidity and surface pressure (O) in the C-univariate (C-uni) and both of the 1-way multivariate data assimilation (1way) and the 1-way multivariate with the variable localization (1way-L). For the multivariate data assimilation without variable localization (Multi), the analysis of wind fields (W) are determined by the coupled error covariance of all the atmospheric variables and CO_2 variables whereas the wind fields from the multivariate with the variable localization (Multi-L) are analyzed without the error correlation with surface CO_2 fluxes (CF).

Table 7. 1. Comparison of all methods: C-univariate (C-uni), Multivariate (Multi), Multivariate with a variable localization between winds and surface CO_2 fluxes (Multi-L), 1-way multivariate (1way), 1-way multivariate with a variable localization (1way-L). W indicates the wind fields, and O includes temperature, specific humidity, and surface pressure, C is the variable of atmospheric CO_2 and CF the surface CO_2 fluxes. The contents of the table present the error information used for the analysis of each variable (W, O, C, CF) in each method (C-uni, Multi, Multi-L, 1way, 1way-L).

	C-uni	Multi	Multi-L	1way	1way-L
W (U,V)	WO	WOCCF	WOC	WO	WO
O (T, q, Ps)	WO	WOCCF	WOC	WO	WO
C (atm CO_2)	CCF	WOCCF	WOCCF	WCCF	WCCF
CF (sfc CO_2 flux)	CCF	WOCCF	CCF	WCCF	CCF

With the experiment of Multi-L and 1way-L shown in Table 7.1, we could see significant improvement of surface CO_2 flux analysis compared to the old multivariate analyses (Multi and 1way). The new one-way multivariate data assimilation (1way-L) thus results in the best performance we were able to obtain in estimating surface CO_2 fluxes compared to the C-univariate and the new multivariate data assimilations. A remaining problem is that surface CO_2 flux field in the imperfect model experiments have "hot spots" over the land in the Southern Hemisphere which are too strong compared to the true fluxes, indicating the presence of a subtle "local EnKF divergence". We address a possible remedy as part of our future work.

7.5. Future plans

Since we do not have these "hot spots" in the perfect model case, they can be assumed to be caused by the model bias of atmospheric CO_2, which we did not correct in the imperfect model set-up. This was confirmed by performing a bias correction for atmospheric CO_2 starting from the nature run and comparing the 6 hour forecasts with it as described in Chapter 4. For the atmospheric variables we know that Reanalyses are accurate enough to capture model biases, so this technique can be applied in the "real world", but we cannot assume that we will have available a Reanalysis that contains atmospheric CO_2 with enough accuracy to use as "nature" and correct the CO_2 bias. Nevertheless, the results confirmed the absence of the "hot spots" when we corrected the bias of the model estimated from the nature run, although in reality we do not have any available dataset on the atmospheric CO_2 to correct the model bias through the low-dimensional correction method (Figure 7.1). Thus, another method for correcting bias of atmospheric CO_2 will be tested and applied. The basic idea of the bias correction is similar to the low-dimensional correction. First, the analysis assimilating the atmospheric CO_2 observation will be done without a bias correction of atmospheric CO_2 as we have done here. Next, we will make 6-hour forecasts which starts from the analysis obtained from the previous step. Then, the analysis increment between the analysis and the forecast will give information about the model bias of atmospheric CO_2. This is a possible way to estimate the model bias of atmospheric CO_2 as long as we have the observations, so this can be applicable to the real case.

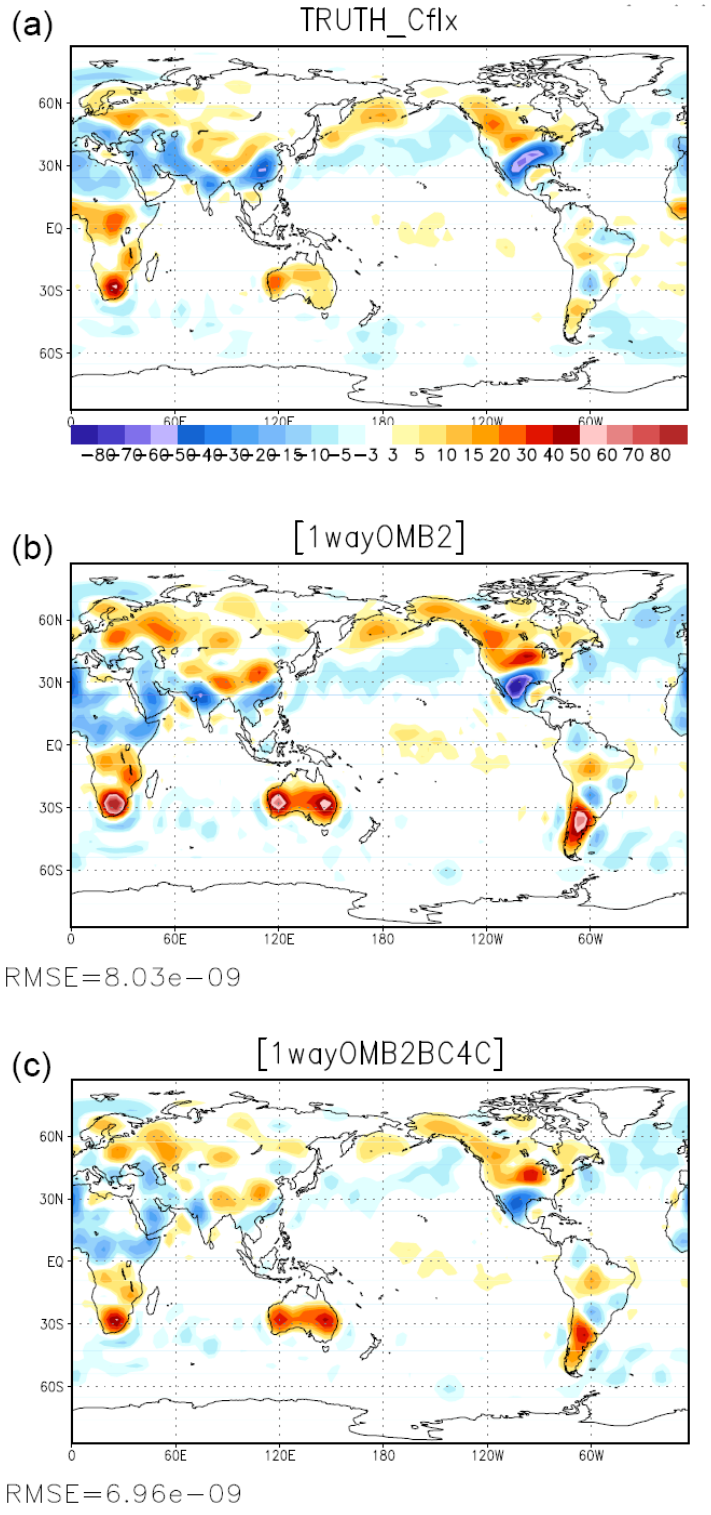

Figure 7. 1. Surface CO_2 fluxes: (a) true state, (b) one-way multivariate data assimilation with an adaptive inflation of the OMB^2 method but no bias correction for atmospheric CO_2, and (c) same as (b) but with a bias correction for atmospheric CO_2 using the low-dimensional method.

This work has been done under the assumption that we have information on atmospheric CO_2 concentrations near the surface. However, it is not yet clear the extent that this we will become true. OCO was going to provide the column-integrated dry air mole fraction data, although the instrument had high sensitivity near the surface. The Japanese satellite GOSAT will also produce the column-based data of CO_2 concentration, and it also has absorption bands sensitive to the atmospheric CO_2 near the surface. Thus, if we have the column data of atmospheric CO_2, then we will calculate the observation increment using the averaging kernel of the dataset. Let a model forecast be $\mathbf{x}^b$ (a CO_2 vertical profile), then

$$\mathbf{y}^b = h(\mathbf{x}^b) = \mathbf{A}^T(\mathbf{H}\mathbf{x}^b) = \sum_{i=1}^{k} a_i(\mathbf{H}\mathbf{x}_i^b) \qquad (7.1)$$

where k is the number vertical levels, $\mathbf{H}$ the spatial interpolation operator, $\mathbf{y}^b$ the model predicted CO_2 column mixing ratio, $\mathbf{A}$ the averaging kernel, and a_i the element of $\mathbf{A}$ at *i-th* vertical level. That is, the observation operator, h, interpolates $\mathbf{x}^b$ to the observation location and also calculates the model forecast as a weighted column CO_2 based on CO_2 profile according to the averaging kernel which is normalized (the sum of $\mathbf{A}$ is equal to unity). We can localize the column observation increment to *i-th* vertical level by the *i-th* averaging kernel element a_i as follows:

$$\Delta\mathbf{y}_i^o = a_i \times (\mathbf{y}^o - \mathbf{y}^b) \qquad (7.2)$$

And the *j-th* ensemble forecast column CO_2 to the *i-th* vertical level by a_i as well:

$$\Delta\mathbf{y}_{j,i}^b = a_i \times \mathbf{y}_j^b \qquad (7.3)$$

Then, we substitute the observation increment of CO_2 (Equation 7.2) in $\Delta \mathbf{y}_i^o = \mathbf{y}^o - \overline{\mathbf{y}}^b$ of Equation (1.4), $\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{X}^b \widetilde{\mathbf{K}} (\mathbf{y}^o - \overline{\mathbf{y}}^b)$, for each vertical layer. The Kalman gain matrix, $\widetilde{\mathbf{K}}$, is calculated as a function of the observation and the forecast error statistics with Equation (1.5), and hence we determine the analysis, $\overline{\mathbf{x}}^a$ in Equation (1.4). In addition, Equations (1.6) and (1.7) provide an estimate of analysis uncertainty. In this we can distribute the observation increment of CO_2 to each vertical layer and assimilate it in LETKF framework. Indeed, this method has been already applied to a realistic system, LETKF/CAM3.5, when assimilating AIRS CO_2 observation and preliminary results are promising (Liu et al., 2009).

This work has been done to test the performance of LETKF on the carbon cycle data assimilation before applying it to a realistic system, LETKF/CAM3.5 using real CO_2 and atmospheric observations. Thus, the insights learned from this study will guide the state-of-art data assimilation system for the carbon cycle and we plan to participate in this process and contribute to solving the problems that will undoubtedly arise when dealing with real observations.

Bibliography

Anderson, J. L. and S.L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127,** 2741–2758.

Andres, R. J., Marland G, Fung I., and Matthews E., 1996: A 1° x 1° distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950-1990, *Global Biogeochem. Cycles*, **10,** 419-429.

Baek, S.-J., B.R. Hunt, E. Kalnay, E. Ott, I. Szunyogh, 2006: Local ensemble Kalman filtering in the presence of model bias, *Tellus*, **58A**, 293-306.

Bacastow, B. R., 1976: Modulation of atmospheric carbon dioxide by the Southern Oscillation. *Nature*, **261**, 116–118.

Baker, D. F., S. C. Doney, and D. S. Schimel, 2006: Variational data assimilation for atmospheric CO_2, *Tellus*, **58B**, 359-365

Barnola, J.-M. 1999. Status of the atmospheric CO_2 reconstruction from ice core analyses. *Tellus* **51B**, 151–155.

Bishop, C. H., Etherton, B. J., and Majumdar, S., 2001: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Mon. Wea. Rev.*, **129**, 420-436.

Bousquet, P., P. Peylin, P. Ciais, C. Le Quere, P. Friedlingstein, and P. P. Tans, 2000: Regional changes in carbon dioxide fluxes of land and oceans since 1980. *Science*, **290,** 1342-1346.

Braswell, B., D. Schimel, E. Linder, and B. Moore, 1997: The response of global terrestrial ecosystems to interannual temperature variability. *Science*, **278**, 870–872.

Buchwitz, M., R. de Beek, J. P. Burrows, H. Bovensmann, T. Warneke, J. Notholt, J. F. Meirink, A. P. H. Goede, P. Bergamaschi, S. Körner, M. Heimann, J.-F. Müller, and A. Schulz, 2005: Atmospheric methane and carbon dioxide from SCIAMACHY satellite data: Initial comparison with chemistry and transport models, *Atmos. Chem. Phys*. **5**, 941-962.

Chahine, M. T., L. Chen, P. Dimotakis, X. Jiang, Q. Li, E. T. Olsen, T. Pagano, J. Randerson, and Y. L. Yung (2008), Satellite remote sounding of mid-tropospheric CO_2, *Geophys. Res. Lett.,* **35**, L17807, doi:10.1029/2008 GL035022.

Crisp, D., R. M. Atlas, F.-M.Breon, L. R. Brown, J. P. Burrows, P. Ciais, B. J. Connor, S. C. Doney, I. Y. Fung, D. J. Jacob, E. Miller D. O'Brien, S. Pawson, J. T. Randerson, P. Rayner, R. J. Salawitch, S. P. Sander, B. Sen, G. L. Stephens, P. P. Tans, G. C. Toon, P. O. Wennberg, S. C. Wofsy, Y. L. Yung, Z. Kuang, B. Chudasama, G. Sprague, B. Weiss, R. Pollock, D. Kenyon, and S. Schroll, 2004: The Obiting Carbon Observatory (OCO) mission. *Advances in Space Research*, **34**, 700-709

Danforth, C.M., E. Kalnay and T. Miyoshi, 2007, Estimating and Correcting Global Weather Model Error. *Mon. Weather Rev.*, **135**, 281-299.

DeFries, R. S., R. A. Houghton, M. C. Hansen, C. B. Field, D. Skole, and J. Townshend, 2002: Carbon emissions from tropical deforestation and regrowth

136 based on satellite observations for the 1980s and 1990s. *Proc. Natl. Acad. Sci.*, **99**(22), 14,256–14,261.

Desroziers G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385-3396.

Enting, I. 2002. Inverse Problems in Atmospheric Constituent Transport. Cambridge University Press, Cambridge, U. K.

Feely, R.A., J. Boutin, C.E. Cosca, Y. Dandonneau, J. Etcheto, H.Y. Inoue, M. Ishii, C.L. Quere, D.J. Machey, M. McPhaden, N. Metzl, A. Poisson, and R. Wanninkhof, 2002: Seasonal and interannual variability of CO_2 in the equatorial Pacific. *Deep- Sea Research Part II-Topical Studies in Oceanography*, **49**, 2443–2469.

Feely, R. A., R. Wanninkhof, W. McGillis, M.-E. Carr, and C. E. Cosca, 2004: Effects of wind speed and gas exchange parameterizations on the air-sea CO_2 fluxes in the equatorial Pacific Ocean, *J. Geophys. Res.*, **109**, C08S03, doi:10.1029/2003JC001896.

Feng, L., Palmer, P. I., Bösch, H., and Dance, S. , 2008: Estimating surface CO_2 fluxes from space-borne CO_2 dry air mole fraction observations using an ensemble Kalman Filter, *Atmos. Chem. Phys. Discuss.*, **8**, 19917-19955.

Friedlingstein, P., P. Cox, R. Betts, L. Bopp, W. von Bloh, V. Brovkin, S. Doney, M. Eby, I. Fung, B. Govindasamy, J. John, C. Jones, F. Joos, T. Kato, M. Kawamiya, W. Knorr, K. Lindsay, H. D. Matthews, T. Raddatz, P. Rayner, C. Reick, E. Roeckner, K.-G. Schnitzler, R. Schnur, K. Strassmann, S. Thompson,

A. J. Weaver, C. Yoshikawa, and N. Zeng, 2006: Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison, *J. Clim.* **19(14)**, 3337–3353.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723-757.

GLOBALVIEW- CO_2, Cooperative Atmospheric Data Integration Project - Carbon Dioxide, CD-ROM, NOAA CMDL, Boulder, Colorado, 2000.

GLOBE Task Team and others (Hastings, David A., Paula K. Dunbar, Gerald M. Elphingstone, Mark Bootz, Hiroshi Murakami, Hiroshi Maruyama, Hiroshi Masaharu, Peter Holland, John Payne, Nevin A. Bryant, Thomas L. Logan, J.-P. Muller, Gunter Schreier, and John S. MacDonald), eds., 1999: The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0. National Oceanic and Atmospheric Administration, National Geophysical Data Center, 325 Broadway, Boulder, Colorado 80305-3328, U.S.A. Digital data base on the World Wide Web (URL: http://www.ngdc.noaa.gov/mgg/topo/globe.html) and CD-ROMs.

Gruber, N., et al., 2009: Oceanic Sources, sinks, and transport of atmospheric CO_2, *Global Biogeochem. Cycles*, **23**, GB1005, doi:10.1029/2008GB003349.

Gurney, K., R. Law, A. Denning, P. Rayner, D. Baker, P. Bousquet, L. Bruhwiler, Y. Chen, P. Ciais, S. Fan, I. Fung, M. Gloor, M. Heimann, K. Higuchi, J. John, T. Maki, S. Maksyutov, K. Masarie, P. Peylin, M. Prather, B. Pak, J. Randerson, J. Sarmiento, S. Taguchi, T. Takahashi, and C. Yuen, 2002: Towards robust

regional estimates of CO_2 sources and sinks using atmospheric transport models. *Nature*, **415**, 626–630.

Gurney, K. R., R. M. Law, A. S. Denning, P. J. Rayner, B. C. Pak, D. Baker, P. Bousquet, L. Bruhwiler, Y. H. Chen, P. Ciais, I. Y. Fung, M. Heimann, J. John, T. Maki, S. Maksyutov, P. Peylin, M. Prather, and S. Taguchi, 2004: Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks. *Global Biogeochemical Cycles*, **18(1),** 10.1029/2003GB002111.

Hamazaki, T., 2008: Greenhouse gasses observation from space: overview of TANSO and GOSAT (http://www.icsoconference2008.com/cd/pdf/S20%20-%20Spectrometers% 20-%20 Hamazaki.pdf)

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Mon. Wea. Rev.*, **129**, 2776-2789.

Hollingsworth, A., R.J. Engelen, C. Textor, A. Benedetti, O. Boucher, F. Chevallier, A. Dethof, H. Elbern, H. Eskes, J. Flemming, C. Granier, J.W. Kaiser, J.J. Morcrette, P. Rayner, V.H. Peuch, L. Rouil, M.G. Schultz, A.J. Simmons, and The GEMS Consortium, 2008: Toward a Monitoring and Forecasting System For Atmospheric Composition: The GEMS Project. *Bull. Amer. Meteor. Soc.*, **89**, 1147–1164.

Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, C. A. Johnson, 2001: Climate Change 2001: The Scientific Basis, *Cambridge U. Press, New York*

Houtekamer, P. L. and Herschel L. Mitchell. 2001: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. Monthly Weather Review: Vol. 129, pp. 123–137.

Hunt, B. R., 2005: An efficient implementation of the Local Ensemble Kalman Filter, http://arxiv.org/abs/physics/0511236.

Hunt, B. R., E. Kostelich, and I. Szunyogh 2007: Efficient Data Assimilation for Spatiotemporal Chaos: a Local Ensemble Transform Kalman Filter, *Physica D*, **230**, 112-126.

Jones, C., M. Collins, P. Cox, and S. Spall, 2001: The carbon cycle response to ENSO: A coupled climate-carbon cycle model study. *J. Climate*, **14**, 4113–4129.

Kalnay, 2003: Atmospheric modeling, data assimilation and predictability, Cambridge University Press, 341 pp.

Keeling, C. D., and R. Revelle, 1985: Effects of ENSO on the atmospheric content of CO_2. *Meteroritics*, **20**, 437–450.

Keeling, C. D., and T. P. Whorf, in Trends: A compendium of data on global change, Carbon dioxide information analysis center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee (2000) (http://cdiac.esd.ornl.gov/ ndps/ndp001.html)

Knorr, W., 2000: Annual and interannual CO_2 exchanges of the terrestrial biosphere: Process-based simulations and uncertainties, *Global Ecol. Biogeogr*., **9(3)**, 225-252.

Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 523-533.

Liu, J., 2007: Applications of the LETKF to Adaptive Observations, Analysis Sensitivity, Observation Impact and the Assimilation of Moisture, Ph.D Thesis, University of Maryland

Liu, J., E. Kalnay, I. Fung, 2009: AIRS CO_2 data assimilation with ensemble kalman filter: preliminary results, *AIRS science team meeting*, Pasadena, CA, May 4-7.

Maddy, E. S., C. D. Barnet, M. Goldberg, C. Sweeney, and X. Liu, 2008: CO_2 retrievals from the Atmospheric Infrared Sounder: Methodology and validation, *J. Geophys. Res.*, **113**, D11301, doi:10.1029/2007JD009402

Maksyutov, S., Kadygrov, N., Nakatsuka, Y., Patra, P. K., Nakazawa, T., Yokota, T., and Inoue, G. , 2008: Projected impact of the GOSAT observations on regional CO_2 flux estimations as a function of total retrieval error, *Journal of Remote Sensing Society of Japan*, **28**, 190-197.

Manning, A.C., and R.F. Keeling, 2006: Global oceanic and land biotic carbon sinks from the Scripps atmospheric oxygen flask sampling network, *Tellus*, **58B**, 95-116.

Mikaloff Fletcher, S. E., et al., 2006: Inverse estimates of anthropogenic CO2 uptake, transport, and storage by the ocean, *Global Biogeochem. Cycles*, **20**, GB2002, doi:10.1029/2005GB002530.

Mikaloff Fletcher, S. E., et al., 2007: Inverse estimate of oceanic sources and sinks of natural CO2 and the implied oceanic carbon transport, *Global Biogeochem. Cycles*, **21**, GB1010, doi:10.1029/2006GB002751.

Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments. *Climate Dyn.*, **20**, 175-191.

Neale, R.B., J.H. Richter, and M. Jochum, 2008: The Impact of Convection on ENSO: From a Delayed Oscillator to a Series of Events. *J. Climate*, **21**, 5904-5924.

Pak, B.C., and M.J. Prather, 2001: CO2 source inversions using satellite observations of the upper troposphere, *Geophys. Res. Lett.*, **28**, 4571-4574.

Peltier, W. R., 1994: Ice Age Paleotopography, *Science*, **265**, pp. 195-201.

Peters, W., J. B. Miller, J. Whitaker, A. S. Denning, A. Hirsch, M. C. Krol, D. Zupanski, L. Bruhwiler, and P.P. Tans, 2005: An ensemble data assimilation system to estimate CO_2 surface fluxes from atmospheric trace gas observations, *J. Geophys. Res.*, **110**, D24304, doi:10.1029/2005JD006157.

Qian, H. 2008: Variability of terrestrial carbon cycle and its interaction with climate under global warming, PhD Thesis at the University of Maryland.

Randerson, J. T., M. V., Thompson, T. J. Conway, I. Y. Fung, and C. B. Field, 1997: The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide, *Global Biogeochem. Cycles*, **11(4),** 535-560.

Rayner, P., R. Law, and R. Dargaville, 1999: The relationship between tropical CO_2 fluxes and the El Niño-Southern Oscillation. *Geophys. Res. Lett.*, **26**, 493–496.

Rödenbeck, C., S., Houweling, M. Gloor, M. Heimann, 2003: CO_2 flux history 1982-2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmos. Chem. Phys.,* **3**, 1919-1964.

Sarmiento, J.L., N. Gruber, 2002: Sinks for anthropogenic carbon, *Physics Today*, **August 2002**, pp. 30-36.

Stockli, R., D. M. Lawrence, G.-Y. Niu, K. W. Oleson, P. E. Thornton, Z.-L. Yang, G. B. Bonan, A. S. Denning, and S. W. Running, 2008: Use of FLUXNET in the Community Land Model development, *J. Geophys. Res.*, **113**, G01025, doi:10.1029/2007JG000562.

Takahashi, T., S. C. Sutherland, C. Sweeney, A. Poisson, N. Metzl, B. Tilbrook, N. Bates, R. Wanninkhof, R. A. Feely, C. Sabine, J. Olafsson, Y. Nojiri, 2002: Global sea-air CO_2 flux based on climatological surface ocean pCO_2, and seasonal biological and temperature effects, *Deep-Sea Research II*, **49**, 1601-1622.

Tans,P.P., I.Y. Fung, T. Takahashi, 1990: Observational constraints on the global atmospheric CO2 buget, *Science*, **247**, 1431-1438.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations, *Mon. Wea. Rev.,* **130(7),** 1913-1924.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review,* **132**, 1238-1253.

Zeng, N., A. Mariotti, and P. Wetzel, 2005: Terrestrial mechanisms of interannual CO_2 variability, *Global Biogeochemical Cycles*, **19**, GB1016, doi:10.1029/2004GB002273

Zeng, N., J. D. Neelin, and C. Chou, 2000: A quasi-equilibrium tropical circulation model-Implementation and simulation, *J. Atmos. Sci*., **57**, 1767-1796

Zeng, N., J-H Yoon, A. Mariotti, and S. Swenson, 2008: Long-term soil moisture variability from a new P-E water budget method. *J. Climate*, **15**, 248-265.

Zupanski, M., S. J. Fletcher, I. M. Navon, B. Uzunoglu, R. P. Heikes, D. A. Randall, T. D. Ringlee and D. Daescu, 2006: Initiation of ensemble data assimilation. *Tellus*, **58A**, 159-170.