

# **Calibrating an earth system model using the adjoint method**

Dissertation zur Erlangung des Doktorgrades  
an der Fakultät für Mathematik, Informatik und Naturwissenschaften  
Fachbereich Geowissenschaften  
der Universität Hamburg  
vorgelegt von Guokun Lyu  
Hamburg, 2017

Tag der Disputation: 05.Dec.2017

Folgende Gutachter empfehlen die Annahme der Dissertation:

Prof. Dr. Detlef Stammer

Dr. Armin Koehl

## Contents

<b>Abstract .....</b>	<b>I</b>
<b>Zusammenfassung .....</b>	<b>III</b>
<b>Acknowledgement .....</b>	<b>V</b>
<b>1 Introduction .....</b>	<b>1</b>
1.1 Climate model tuning.....	1
1.2 Parameter optimization methods.....	2
1.3 Objectives .....	3
1.4 Structure of this thesis.....	4
<b>2 Background.....</b>	<b>5</b>
2.1 The Bayesian formulation.....	5
2.2 Stochastic Bayesian methods.....	6
2.3 Optimization algorithms .....	7
2.4 Data assimilation methods .....	8
2.4.1 <i>Ensemble Kalman Filters</i> .....	8
2.4.2 <i>The adjoint method</i> .....	9
2.5 Summary .....	10
<b>3 Methodology .....</b>	<b>11</b>
3.1 Mathematical formulation of variational data assimilation .....	11
3.2 The adjoint method .....	12
3.2.1 <i>The tangent linear model and its adjoint</i> .....	12
3.2.2 <i>The minimization algorithm</i> .....	14
3.2.3 <i>The optimization process</i> .....	15
3.3 Limitations of the adjoint method.....	16
3.3.1 <i>Behaviour of the cost function</i> .....	17
3.3.2 <i>Behaviour of the adjoint sensitivities</i> .....	18
3.4 Extending the feasible assimilation window.....	19
3.4.1 <i>The dynamical state and parameter estimation method</i> .....	20
3.4.2 <i>The modified DSPE method</i> .....	22
<b>4 Assimilating experiments with the Lorenz system .....</b>	<b>25</b>
4.1 Case studies.....	25
4.1.1 <i>Perfect observations</i> .....	25
4.1.2 <i>Noisy observations</i> .....	27
4.2 Dependence on synchronization efficiency .....	28

4.2.1 <i>Influence of observation noise</i> .....	29
4.2.2 <i>Experiments with different nudging variables</i> .....	31
4.2.3 <i>Influence of observation frequency</i> .....	33
4.3 Discussions and conclusions.....	36
<b>5 The Climate model and configurations .....</b>	<b>38</b>
5.1 The model and its configurations.....	38
5.2 Description of physical parameterizations.....	41
5.2.1 <i>Rayleigh friction</i> .....	42
5.2.2 <i>Horizontal diffusion</i> .....	42
5.2.3 <i>Vertical diffusion parameterization</i> .....	42
5.2.4 <i>Radiation parameterization</i> .....	43
5.3 Implementation of nudging terms .....	44
5.4 Summary .....	44
<b>6 Assimilation experiments with PlaSim .....</b>	<b>45</b>
6.1 The cost function.....	45
6.2 Modifying the TLM and its adjoint .....	45
6.3 Assimilating experiments with the “minimal” configuration .....	49
6.3.1 <i>The nudging strength for synchronization</i> .....	49
6.3.2 <i>Assimilation experiments</i> .....	52
6.4 Assimilating experiments with the “maximal” configuration .....	53
6.4.1 <i>The nudging strength for synchronization</i> .....	55
6.4.2 <i>Assimilation experiments</i> .....	60
6.4 Discussions and conclusions.....	63
<b>7 Assimilating the ERA-Interim data.....</b>	<b>65</b>
7.1 Experiment setups.....	65
7.2 Results for the first group assimilation experiments.....	67
7.2.1 <i>Convergence of the minimization</i> .....	67
7.2.2 <i>Evaluation of the optimization</i> .....	69
7.2.3 <i>Detailed effects of each parameter</i> .....	73
7.2.4 <i>Assimilation of the temperature observations</i> .....	83
7.2.5 <i>Summary</i> .....	84
7.3 Results for the second group assimilating experiments .....	85
7.3.1 <i>Convergence of the minimization</i> .....	85
7.3.2 <i>The air temperature improvement</i> .....	86
7.3.3 <i>The radiative fluxes improvement</i> .....	88

7.4 Evaluation with the free model .....	93
7.4.1 <i>The Radiative fluxes</i> .....	93
7.4.2 <i>Temperature and zonal wind</i> .....	96
7.4.3 <i>Specific humidity, precipitation and cloud cover</i> .....	98
7.5 Summary .....	100
<b>8 Conclusions and Future works.....</b>	<b>103</b>
8.1 Overview of Conclusion .....	103
8.2 Outlook .....	106
<b>Table Captions.....</b>	<b>113</b>
<b>Reference.....</b>	<b>114</b>

# Abstract

This thesis investigates the potential of the adjoint method for calibrating a climate model. The adjoint method is applied to optimize process parameters on climate timescales to reduce model biases. The difficulty that must be overcome is the limited assimilation window in the adjoint method. Extending the assimilation window longer than the characteristic period of the fastest growing mode, will lead to the occurrence of secondary minima accompanied by an exponential increase of the adjoint sensitivities, and the gradient-descent minimization algorithm is likely trapped into local minima. With a long assimilation window such as climate timescales, the adjoint model cannot provide useful gradients for the optimization. To overcome the limited assimilation window problem, synchronization which is implemented as nudging technique is exploited to regularize the fast-growing modes of the nonlinear system and hence extend the feasible assimilation window for parameter estimation.

Firstly, the performance of this method was investigated based on Lorenz (1963) model. It was shown that: by using a finite nudging coefficient which is strong enough to push the positive Lyapunov exponents to negative values, the feasible assimilation window can be extended arbitrary and the control parameter can be efficiently and reasonably retrieved. Performance of this method depends on synchronization efficiency which is influenced by observation noise, observation frequency, variables chosen for nudging and nudging strength. With noisy and sparse observations, an optimal nudging coefficient which best recovers true signal can be predefined and benefits the parameter estimation.

Secondly, this method was applied to an intermediate earth simulation model, the Planet Simulator (PlaSim). I closely examined the usefulness of the adjoint model generated by an automatic differentiation tool TAF. Then identical twin experiments were performed with two different configurations, with and without moisture parameterizations (the ‘maximal’ and ‘minimal’ configurations, respectively). The optimization successfully retrieved the default values of the control parameters for both the two configurations with assimilation window of 2-month and 1-year.

At last, the ‘maximal’ configuration was used to optimize process parameters by assimilating the ERA-Interim data. A number of assimilation experiments using 4,7,16 control parameters and using different observations in the cost function were conducted. The contributions of each parameter to the model state variables were studied in detail. By optimizing two parameters controlling absorptivity (longwave) of clouds and water vapor, the global mean bias of net long wave radiation at the surface and at the top of the atmosphere can be significantly reduced. The global mean bias of short wave radiation at the surface and at the top of the atmosphere can be efficiently reduced by optimizing parameters tuning cloud optical properties. The air temperature is also considerably improved. Then, the estimated parameters were tested with the free model (without nudging terms). The improvements in the radiative fluxes and the air temperature are similar to that in the assimilation experiments which further validate the usefulness of the method. Other model states such as convective

precipitation and surface latent heat flux show both improvement and deterioration. However, the specific humidity is hardly improved which is likely due to model deficiency.

This study demonstrates that by using synchronization, the adjoint method can be applied to estimate process parameters on climate timescales efficiently. The method overcomes difficulties of parameter estimation in chaotic models and provides a promising way for tuning process parameters in coupled climate models.

# Zusammenfassung

Diese Arbeit untersucht das Potenzial der adjungierten Methode zur Verbesserung eines Klimamodells. Das maximale Assimilationsfenster in der adjungierten Methode ist durch die Vorhersagbarkeit des nichtlinearen Systems begrenzt. Eine weitere Erweiterung des Assimilationsfensters führt zum Auftreten von mehreren Minima in der objektiven Funktion, die die Abweichung zwischen Model und Daten misst, begleitet von einer exponentiellen Zunahme der Gradienten. Als Folge steigt die Wahrscheinlichkeit, dass die Minimierung in einem lokalen Minimum stecken bleibt. Um die Beschränkung des Assimilationsfensters zu überwinden, wird Synchronisation mit den Daten ausgenutzt, die die schnell wachsenden Modi dämpft und das nichtlineare System regularisieren kann.

Zunächst wurde diese Methode anhand des Lorenz (1963) Modells untersucht. Es wurde gezeigt, dass durch die Verwendung eines festen Nudging-Koeffizienten, der groß genug ist, um positiven Lyapunov-Exponenten auf negative Werte zu drücken, das Assimilationsfenster, in dem der Parameter effizient rekonstruiert werden kann, beliebig erweitert werden kann. Diese Methode hängt von Beobachtungsfehlern, Beobachtungshäufigkeit, den Variablen, die für die Kopplung gewählt wurden, und deren Kopplungsstärke ab, da diese Faktoren zur Synchronisationseffizienz beitragen. Bei fehlerhaften und spärlichen Beobachtungen ist der optimale Kopplungskoeffizient dadurch ausgezeichnet, dass das fehlerfreie Signal durch Synchronisation am besten angenähert wird. Dieser Wert ist auch optimal für die Parameterschätzung, da der Einfluss der Fehler auf die geschätzten Parameterwerte minimal wird.

Im zweiten Teil wurde diese Methode auf ein Erdsystem-Simulationsmodell, den Planet Simulator (PlaSim), angewendet. Wir untersuchten die Nützlichkeit des adjungierten Modells, das durch das automatische Differenzierwerkzeug TAF erzeugt wurde, für die Parameteroptimierung. Es wurden identische Zwillingsexperimente mit zwei verschiedenen Konfigurationen durchgeführt; mit und ohne Feuchtigkeitsparametrierungen ("Maximale" und "Minimale" -Konfiguration). Die Optimierung hat die Standardwerte der Parameter für beide Konfigurationen mit einem Assimilationsfenster von 1 Jahr erfolgreich rekonstruiert.

Schließlich wurde die "Maximale" Konfiguration verwendet, um Prozessparameter zu optimieren, indem ERA-Interim-Daten assimiliert werden. Eine Reihe von Assimilationsexperimenten unter Verwendung von 4, 7 oder 16 Kontrollparametern und unter Verwendung verschiedener Beobachtungen in der objektiven Funktion wurden durchgeführt, um die Leistungsfähigkeit dieser Methode zu bewerten. Die Beiträge der einzelnen Parameter zu verschiedenen Modelvariablen wurden im Detail untersucht. Durch die Optimierung von 2 Parametern, die die Absorptionsfähigkeit (Langwellen) von Wolken und Wasserdampf kontrollieren, kann die globale mittlere Vorspannung der Netto-Langwellenstrahlung an der Oberfläche und an der Oberseite der Atmosphäre deutlich reduziert werden. Die global gemittelten Fehler der Kurzwellenstrahlung an der Oberfläche und an der Oberseite der Atmosphäre konnten effizient reduziert werden, indem die Parameter optimiert wurden, die die optischen Eigenschaften der Atmosphäre bestimmen. Auch die Lufttemperatur wurde deutlich verbessert. Dann wurden die geschätzten Parameter mit dem freien Modell (ohne

Synchronisation) getestet. Die Verbesserungen für die Strahlungsflüsse und die Temperatur sind ähnlich denen der Assimilationsexperimente, womit die Nützlichkeit des Verfahrens weiter validiert worden konnte. Andere Modellzustände wie konvektiver Niederschlag und latenter Oberflächenwärmefluss wurden ebenfalls verbessert. Allerdings wurde die spezifische Feuchtigkeit kaum verbessert.

Diese Studie zeigt, dass durch Synchronisation die adjungierte Methode angewendet werden kann, um auf klimatologischen Zeitskalen Prozessparameter effizient abzuschätzen. Die Methode überwindet Schwierigkeiten bei der Parameterschätzung in chaotischen Modellen und bietet eine vielversprechende Möglichkeit, Prozessparameter in gekoppelten Klimamodellen abzustimmen.

# **Acknowledgement**

I would like to show my deep gratitude to my scientific supervisor Prof. Detlef Stammer for providing me an opportunity to study in his group and providing me funding over the past year. I would like to thank my co-supervisor Dr. Armin Koehl for motivating this research, for his advices and useful discussions. They also generously provided comments on my drafts.

I would also like to thank Dr. Ion Matei for helping me use the PlaSim model and sharing me his knowledge about the model setup.

Special thanks to my families for their financial and mental support. Many, many thanks to my wife Xiangxin, for her encouragement and support every step of the way.

Last but not the least, I gratefully acknowledge the financial support from the China Scholarship Council (CSC) from 01.10.2013-30.09.2016.

# Chapter 1

## Introduction

### 1.1 Climate model tuning

State-of-the-art climate models are widely used both to project the future climate change and understand the dynamics of the atmosphere and ocean system. Climate models solve the primitive equations which govern the atmosphere and ocean system, using different numerical methods that are specified on a specific spatial grid. Due to limited computational resources, the spatial resolution of climate models is usually very coarse, typically ranging between 25-300km for global models (Taylor et al., 2012). Many processes such as cloud formation, radiative transfer, and turbulence, occur at much smaller scales and thus cannot be explicitly resolved by the numerical models. In such a case, parameterizations are used to approximate the statistical impacts of the sub-grid-process on the large-scale flow regarding the large-scale flow itself. The details of these parameterizations and values of process parameters involved are essential to determine the model's climatology (Murphy et al. 2004). But because parameterization details, as well as the dependence of their amplitudes on the simulated large-scale flow fields, are largely unknown, current studies use process parameters that are constant in time and space. Unfortunately, even these constant values span a large range of plausible values (Järvinen et al., 2010; Louis, 1979; Murphy et al., 2004; Tett et al., 2013). Thus, tuning these uncertain parameters is crucial to improve the climate simulation and affects climate sensitivity (Hourdin et al., 2016).

Traditionally, process parameters are manually tuned in a trial and error approach to reduce the difference between the model-simulated and observed climatologies. Mauritsen et al. (2012) documented how they tuned a global model. Based on a few well-understood cases for which the impact a process parameter on the model simulation is known, they tuned the climate model to best match the radiation balance at the top of the atmosphere (TOA), the global mean temperature, sea ice, clouds and wind fields by adjusting couples of uncertain parameters. Due to the complexity of climate models, tuned process parameters are limited to a very small subspace, and observations that are used as targets are often related to the energy balance of the climate models. A survey on “How do modeling centers tune their models?”<sup>1</sup> shows that parameters related to clouds are most frequently tuned, and the TOA radiation balance and the global mean temperature are commonly used as targets (Hourdin et al., 2016; Mauritsen et al., 2012). This aims at balancing energy received from the sun and energy lost to space by adjusting the representation of clouds. For different purposes, additional targets may also be included in the tuning process such as tropical variability or ocean heat transport in the North Atlantic. Thus, the traditional tuning process is subjective, computationally expensive and labor-intensive. A further complication is that it is usual for a tuning process to be unsuccessful. Even if only a handful of process parameters are considered, their non-local effects can lead to unforeseeable, sometimes detrimental, effects over long integration periods. Additionally, the values of process parameters may depend on the spatial and temporal

---

<sup>1</sup> See <http://dx.doi.org/10.1175/BAMS-D-15-00135.2>

resolution of the model (Tiedtke, 1989). Because of this, the modelling community is looking for alternatives to efficiently and effectively tuning process parameters that always is intended to improve the model simulation. Parameter optimization techniques are such alternatives and have advanced quickly over the past decades.

## 1.2 Parameter optimization methods

Considerable parameter optimization methods have been developed by statistics, engineering and computer science communities and have been applied to optimize process parameters for climate models. Generally, these methods can be categorized into three groups based on Stochastic Bayesian approaches, optimization algorithms and data assimilation techniques.

In Stochastic Bayesian approaches (Järvinen et al., 2010; Jackson et al., 2004), optimized parameters and confidence ranges are estimated from prior probability of parameters and a likelihood function that measures the probability of the observations simulated by the models. These methods require a large number of model integrations and are computationally expensive. For example, Järvinen et al. (2010) estimated 3 of 4 parameters based on the adaptive Markov chain Monte Carlo (MCMC) method which requires 4500 years of model integrations. Also, these methods depend on the choice of the likelihood function.

Optimization algorithms are exploited to find values of a set of parameters that minimize an objective function which measures the distance of the model simulation to a small subset of observations. Some existing methods such as Green's function method (Menemenlis et al., 2005; Stammer and Wunsch, 1996), the simultaneous perturbation stochastic approximation method (SPSA, Spall, 1998) and Gauss-Newton line-search algorithms (Tett et al., 2013; Tett et al., 2017) are investigated with the aim of automatically calibrating climate models. Usually, gradients of the cost function with respect to process parameters are approximated with finite differences of perturbed integrations. Convergence and costs are sensitive to the number of process parameters and optimization algorithms used.

Being an optimization problem, data assimilation would be an alternative as well and can efficiently handle a large number of observations. It has been well addressed for state estimation and has been investigated for parameter estimation. Some studies have estimated the uncertain parameters using an ensemble Kalman Filter (Annan et al., 2005; Evensen, 2009; Schirber et al., 2013; Wu et al., 2012; Zhang, 2011) and achieved the success of different degrees. For the ensemble Kalman filter method, parameters are tuned based on the covariance of the parameters and the model states, which depends on the ensemble number. Larger ensemble numbers may better depict their covariance, but they will also significantly increase computational costs. The adjoint method is another advanced data assimilation method which is widely used to estimate the model states (Dee et al., 2011; Köhl, 2015; Saha et al., 2010; Stammer et al., 2002) and the parameters (Liu et al., 2012). Using the German Estimating the Circulation and Climate of the Ocean (GECCO) synthesis framework, Liu et al. (2012) showed that the adjoint method could be used to efficiently estimate the parameters that control eddy-induced tracer mixing. For atmospheric models, on the other hand, it is challenging to estimate process parameters on climate timescale simulation by using the

adjoint method. The reasons are: (1) some of the parameterization schemes take very complex forms which make it hard to code the tangent linear model and its adjoint; (2) some of the parameterized processes such as convective precipitation are highly nonlinear and may degrade usefulness of the adjoint model; (3) a large assimilation window is required to allow the process parameters to influence the model's climatology while the feasible assimilation window of the adjoint method is limited by predictability of the nonlinear system.

In this study, the adjoint method will be further investigated for calibrating a climate model by optimizing process parameters. This study is motivated by the following two studies: (1) based on CEN Earth System Assimilation Model, Blessing et al. (2014) showed that it is feasible to use the adjoint model automatically generated by the Transform Algorithm of Fortran (TAF, Giering; Kaminski 1998) to estimate process parameters. With an automatic differentiation tool TAF, the adjoint model is generated although with the complex forms of parameterization schemes in this coupled model. This work solves the first problem as mentioned in the last paragraph; (2) Abarbanel et al. (2010) showed the potential of using synchronization to overcome the limited assimilation window problem in the adjoint method. It is feasible to investigate whether the adjoint method can be used to estimate process parameters on climate timescales efficiently.

Further, in previous studies based on the traditional tuning method and optimization algorithms, the observations used are usually global mean values, sometimes zonally averaged values (Järvinen et al., 2010; Jackson et al., 2004; Zhang et al., 2015). Only a small subset of observations is included in the objective functions. Different climate models may be tuned to perform better on constrained variables while degrade other variables. Nowadays, reanalysis datasets (Dee et al., 2011; Saha et al., 2010) provide high resolution and high-frequency model states which better describe the atmosphere system. Making full use of the reanalysis datasets to calibrate climate models is preferable. The adjoint method appears as the most natural choice with computational costs that are substantially independent of the control parameter dimensions (within the same iterations) and can easily handle a plenty of observations.

### 1.3 Objectives

This thesis aims to estimate process parameters on climate timescales using the adjoint method. The Dynamical State and Parameter Estimation (DSPE) method (Abarbanel et al., 2010; Abarbanel et al., 2009) will be exploited to extend the feasible assimilation window to climate timescales and tested for parameter estimation. The following work will be done:

- (1) Testing the performance of the DSPE method with the Lorenz (1963) system. Potential problems with the classical DSPE method will be explored and explained. Subsequently, modifications will be made on this classical DSPE method for parameter estimation. Finally, factors that influence the accuracy of parameter estimation with the modified DSPE method will be comprehensively investigated.
- (2) The modified DSPE method will be applied to the Planet Simulator (PlaSim, Fraedrich et al. 2005b). A ‘maximal’ configuration and a ‘minimal’ configuration which varies in their degrees of nonlinearity are used for testing. With the adjoint

model generated by TAF, identical twin experiments are performed to investigate the viability of parameter estimation on climate timescales with the modified DSPE method.

- (3) Lastly, the ‘maximal’ configuration will be applied to estimate parameters by assimilating ERA-Interim data. A number of assimilation experiments will be conducted. The effects of each parameter to the observations will be analyzed in detail. Then, impacts of the estimated parameters on the free model (without nudging terms) will be evaluated to validate the usefulness of this method.

#### **1.4 Structure of this thesis**

The rest of this study is structured as follows:

Chapter 2 provides the background of parameter optimization problem starting from Bayes’ theorem. Several parameter estimation methods that have been applied to optimize process parameters for calibrating climate models are discussed.

Chapter 3 provides the methodology for this study. The adjoint method is introduced based on the Lorenz (1963) system. The challenge of applying the adjoint method to climate timescales assimilation is presented. The potential of using chaos synchronization (Abarbanel et al., 2010; Abarbanel et al., 2009) to overcome the challenge is discussed, and this method is further modified for parameter estimation.

Chapter 4 investigates the performance of this modified DSPE method based on the Lorenz (1963) system. Assimilation experiments are performed to investigate the dependence of the modified DSPE method on observation noise, observation frequency and other factors.

Chapter 5 gives a detailed description of the climate model, PlaSim, used in this study. The parameterizations and the process parameters to be optimized are also described.

Chapter 6 further applies this modified DSPE method to PlaSim. The usefulness of the adjoint model generated by TAF is closely examined based on finite differences of perturbed integrations. Finally, identical-twin experiments are performed to investigate the performance of the method with this complex earth system model.

Chapter 7 estimates process parameters by assimilating ERA-Interim data (Dee et al., 2011) for an assimilation window of 1 year. Several experiments are performed to test the performance of the modified DSPE method. The influence of estimated parameters on the free model’s climatology is evaluated.

Finally, chapter 8 compiles the conclusions of this study and outlines future work.

# Chapter 2

## Background

Parameter estimation tries to find optimal values of a set of parameters which minimize the difference between model-simulated and observed climatologies. Parameter estimation problem can be coherently described by Bayes' theorem. This chapter will provide a systematic introduction to parameter estimation problem in the framework of Bayes' theorem, and the three categories of optimization methods mentioned above will be derived in this framework. Then, assumptions made in different optimization methods will be discussed, and the potential of using data assimilation techniques for calibrating climate models will be highlighted.

### 2.1 The Bayesian formulation

Bayes' theorem is stated mathematically as equation (2.1):

$$P(m|d_o) = \frac{P(d_o|m)P(m)}{P(d_o)} \quad (2.1)$$

where  $m$  represents a poorly known parameter vector and  $d_o$  represents an observation vector.  $P(m)$  and  $P(d_o)$  are prior probabilities of the parameter vector  $m$  and the observation vector  $d_o$  without regarding the other.  $P(d_o|m)$  is the probability of observing observations  $d_o$  given the parameter vector  $m$  (usually called likelihood functions).  $P(m|d_o)$  gives a conditional probability of the parameter vector  $m$  given the observation vector  $d_o$ . Equation (2.1) outlines a solution to an inverse problem from which one can get the best choice of the parameter vector  $m$  that is consistent with the observation vector  $d_o$ . The uncertainty of the parameter vector  $m$  can also be derived from the posterior probability  $P(m|d_o)$ .

Figure 2.1 gives a simple linear example in which parameter uncertainty takes a Gaussian distribution. The solid blue line in Figure 2.1 shows the prior probability of parameter  $m$  with its mean  $\langle m \rangle = 0$  and uncertainty  $\sigma_m = 2.0$ . The linear equation is  $x_1 = 2*m$ . An observation  $d_o = 5.0$  is given with uncertainty of  $\sigma_{d_o} = 0.5$  and also takes a Gaussian distribution. The posterior probability of  $m$  can be expressed by:

$$P(m|d_o) = \frac{\frac{1}{\sqrt{2\pi}\sigma_{d_o}} \exp\left(-\frac{(x_1 - d_o)^2}{2\sigma_{d_o}^2}\right) \times \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left(-\frac{(m - \langle m \rangle)^2}{2\sigma_m^2}\right)}{\int \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left(-\frac{(m - \langle m \rangle)^2}{2\sigma_m^2}\right) \times \frac{1}{\sqrt{2\pi}\sigma_{d_o}} \exp\left(-\frac{(x_1 - d_o)^2}{2\sigma_{d_o}^2}\right) dx} \quad (2.2)$$

The solid red line and the solid black line in Figure 2.1 represent the likelihood function, and the posterior probability distribution of parameter  $m$  with given observation  $d_o$ . In climate models ( $x_1$  in this example),  $m$  represents model uncertain parameters or the initial state and  $d_o$  represents available observations. With relatively accurate observations, parameter estimation attempts to derive the posterior probability of parameters which enables better simulation of the climate system and uncertainty analysis.

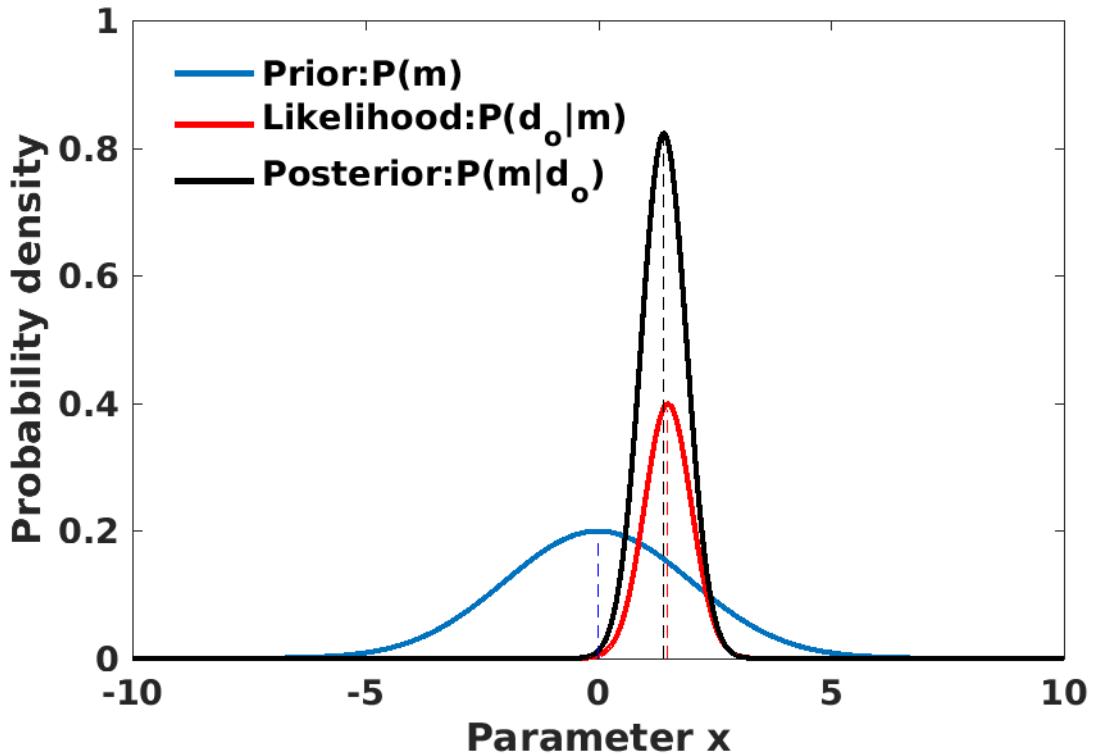


Figure 2.1. A prior probability distribution of the uncertain parameter  $m$  (solid blue line), a likelihood function of observing the observations  $d_o$  given the uncertain parameter  $m$  (solid red line) and the posterior probability distribution estimated based on Bayes' theorem (solid black line). The dash lines represent the maximum probabilities of the three probability density functions.

## 2.2 Stochastic Bayesian methods

In a linear model, the posterior probability density function can be explicitly solved as equation (2.2) and its time evolution can be characterized by its mean value and variance. But in climate models which consist of several nonlinear prognostic equations, explicitly calculating the posterior probability is impossible. This problem can be solved by sampling parameter space with large ensemble members and then calculate an approximated posterior probability.

The prior probability of  $P(m)$  is usually based on our knowledge of the parameter and the climate system (e.g. a uniform distribution or a Gaussian distribution), so

$$P(m) = \sum_{i=1}^N G(m_i) \quad (2.3)$$

where  $G$  is a given probability density function of parameter vector  $m$ ,  $N$  is the total ensemble members and equation (2.1) is approximated with:

$$P(m|d_o) = \sum_{i=1}^N \frac{P(d_o|m_i)G(m_i)}{\sum_{j=1}^N P(d_o|m_j)G(m_j)} \quad (2.4)$$

Under the Gaussian assumption, the likelihood function  $P(d_o|m)$  takes the following form:

$$P(d_0|m_i) = A \exp\left(-\frac{(d_0 - M(m_i))^2}{2\sigma^2}\right) \quad (2.5)$$

and  $M$  is both the model operator and the measurement operator which transform parameter  $m$  information to estimates of the observations  $d_0$ .  $\sigma$  is uncertainty of the observations and  $A$  is a constant that doesn't influence final estimation.

The direct method to solve equation (2.4) involves subdividing parameter space into a number of equally spaced intervals. Then computing the likelihood function  $P(d_0|m)$  for every possible combination of parameters and then evaluating the posterior probability in equation (2.4). The advantage of the direct method is that the full posterior probability can be computed with sufficient small parameter intervals. The disadvantage is the large number of model integrations, and the resolution of the posterior probability is constrained by the interval spacing. For example, if one parameter is divided into 100 equally spaced intervals, the model needs to be integrated 100 times to calculate  $P(m|d_0)$ . Many of these integrations may have very small prior probability, hence don't contribute to the integral in equation (2.4). Moreover, increasing the number of parameters will lead to dramatically increase of model integrations with a rate of  $100^n$  in which  $n$  represents the number of parameters.

To reduce computational costs of this equal sampling method, importance sampling techniques such as Gibbs sampler (Jackson et al., 2004), adaptive Markov chain Monte Carlo method (Järvinen et al., 2010) and multiple very fast simulated annealing (Jackson et al., 2004) are exploited to improve efficiency of calculating the posterior probability of process parameters. Based on given probability distributions of process parameters, these importance sampling techniques will sample more frequently in larger likelihood region. Although with different efficient sampling methods, estimating the posterior probability of process parameters is still computationally expensive and the computation cost may increase dramatically with increasing of parameter dimensions. For examples, in Jackson et al. (2004)'s work, estimating three parameters cost more than 34000 model integration for multiple very fast simulated annealing, and  $10^4$ - $10^6$  integration for Gibbs sampler and grid search method. From the model development perspective, determining the optimal values of parameters that maximize the likelihood function  $P(d_0|m)$  or minimize the difference between the model-simulated and observed climatologies is more attractive.

### 2.3 Optimization algorithms

Optimization algorithms are developed to find values of a set of parameters that minimize an objective function (hereafter, called cost function) which measures the distance between the observations and the model simulations. The cost function usually takes a quadric form as equation (2.6):

$$J(m) = \frac{(d_0 - M(m))^2}{2\sigma^2} \quad (2.6)$$

which can be derived by a simple natural logarithmic transformation of equation (2.5).

Optimization algorithms are originally applied to linear systems. For climate models which are nonlinear, most of the practical estimation methods use linearization assumption in which sensitivity of the model's climatology with respect to process parameter  $m$  is assumed

linear. Then, gradients of the cost function  $J(m)$  with respect to parameter  $m$  can be approximated by finite differences of perturbed integrations as equation (2.7) shows and the cost function  $J(m)$  is recursively reduced until certain criteria are reached.

$$\frac{\partial J}{\partial m} = \frac{J(m+\delta m) - J(m)}{\delta m} \quad (2.7)$$

A common characteristic of optimization algorithms is that they assume linear responses of the cost function to parameters. But different optimization methods may work differently to search the optimal values of process parameters. For example, Tett et al. (2013) use Gaussian-Newton line search algorithms and approximate gradients from finite differences of perturbed integrations to automatically calibrate HadAM3. The Green's function method is another popular optimization method which follows inverse theory and has been successfully applied to calibrate an Ocean General Circulation Model (Menemenlis et al., 2005; Stammer and Wunsch, 1996). Rather than directly reducing the cost function  $J(m)$ , this method maps the observation information back to process parameters via the model Green's function which is computed based on a linearization assumption. For these two methods, the approximate gradients with respect to each parameter need to be evaluated one by one and therefore the computational cost is proportional to the dimension of parameters. The simultaneous perturbation stochastic approximation (SPSA) method is a highly efficient optimization algorithm and has been applied to calibrate an atmosphere general circulation model (Agarwal, 2016). The approximate gradients are evaluated based on only two simultaneous perturbed integrations regardless of the dimension of parameters.

A key advantage of optimization algorithms is the simplicity of implementation compared with the adjoint method or the ensemble Kalman Filter. But a principal drawback of these methods is that the cost function needs to be reasonably linear. Usually, observations are globally averaged to smooth out spatial variance which makes approximate gradients from finite differences useful. Moreover, convergence and the computational cost likely depend on gradient descent algorithms, number of process parameters and nonlinearity of climate models.

## 2.4 Data assimilation methods

Data assimilation techniques have been well addressed for state estimation and have been widely used to produce accurate initial state for forecasting (Schiller and Brassington, 2011). Data assimilation methods also show potential for parameter estimation and several studies have applied these methods to estimate parameters. One of the advantages of data assimilation methods is that they can easily handle a plenty of observations. Here, we introduce two advanced data assimilation methods which have been investigated for parameter estimation.

### 2.4.1 Ensemble Kalman Filters

The ensemble Kalman filer (Evensen, 2009) is a popular data assimilation method due to its easy implementation. The ensemble Kalman Filter bases on least square fitting algorithm and is widely used to blend observations and model initial state for forecasting. By taking

process parameters as model state variables, it is also extended to estimate parameters (Anderson, 2001; Annan et al., 2005; Wu et al., 2012; Zhang, 2011).

The ensemble Kalman Filter uses finite ensemble members to represent the nonlinear propagation of prior probabilities. When observations are available, the Kalman Filter algorithm will update the model state based on the model and the observation uncertainty. The covariance matrix computed from the ensemble members also helps to distribute observed information to unobserved model state and parameters. In this method, the covariance matrix plays a critical role in the analysis process. For example, Schirber et al. (2013) provided a study in which they use an ensemble Kalman Filter to estimate four cloud-related parameters based on ECHAM6 and found no improvements on the model's climatology. One of the possible reasons is that they use a 6-hour update interval and therefore the covariance matrix may only reflect covariance of short-term prediction with parameters. The parameters are tuned to improve short-term forecast rather than the model's climatology. Annan et al. (2005) also attempted to estimate process parameters in an intermediate complexity earth system using an ensemble Kalman Filter. To tune the model's climatology, they used a 10-year update interval. Based on identical twin experiments, they demonstrated that the model's climatology could be tuned by simultaneous estimation of 12 parameters. But applying this method to assimilate real observations needs further investigation. In the ensemble Kalman Filter method, some artificial tricks such as covariance inflation or localization are needed to get good results in a complex geophysical system.

#### **2.4.2 The adjoint method**

The adjoint method (Le-Dimet and Talagrand, 1986; Talagrand and Courtier, 1987), also known as four-dimensional variational data assimilation (4D-Var) in numerical weather prediction community, is a mature data assimilation method. This method belongs to optimization algorithms because this method also tries to find optimal values of the control variables including process parameters and model state that minimize a cost function as equation (2.8). However, the adjoint method is much more complicated and optimal than the optimization algorithms introduced above due to the use of the adjoint model. The adjoint model is exploited to calculate gradients of the cost function with respect to the control variables. It is a popularly used inverse method for reanalyse (Dee et al., 2011; Köhl, 2015; Saha et al., 2010) and parameter estimation (Liu et al., 2012; Zou et al., 1992).

$$J(m) = \frac{(d_0 - M(m))^2}{2\sigma^2} + \frac{(m - m_b)^2}{2\sigma_b^2} \quad (2.8)$$

If parameter  $m$  takes a Gaussian distribution with mean  $m_b$  and uncertainty  $\sigma_b$ , equation (2.8) can be derived by a simple natural logarithmic transformation of equation (2.4). It should be noted that the denominator part of equation (2.4) can be thought of as a normalizing constant and therefore doesn't matter in the adjoint method. The cost function (2.8) is assumed to be quadric, and a minimization algorithm (such as the Quasi-Newton algorithms) is adopted to iteratively reduce the cost function, or equivalently maximize the posterior probability  $P(d_0|m)$ .

In the adjoint method, the second term on the right-hand side of equation (2.8) is usually called a background term. It is critical when there are more control variables than the observations (or the minimization problem is under-determined). But in this study, only a few process parameters will be tuned while a plenty of observations can be used. Also, prior probabilities of process parameters are not clear. Therefore, the background terms will not be included in the cost function.

The adjoint method provides an analysis that is consistent with model dynamics and its computational cost, while substantial, is independent of control variables. Therefore, it has been the method of choice for most major numerical weather prediction centers and widely used for reanalyse. It also shows great potential for parameter estimation. But constructing the adjoint model is a long-term project along with the development of the forward models, and the feasible assimilation window in this method is limited by the predictability of nonlinear system (will be introduced in chapter 3). Techniques are required to extend the feasible assimilation window for the adjoint method.

## 2.5 Summary

In this part, parameter estimation problem is formulated based on Bayes' theorem. The three categories of parameter estimation methods are briefly discussed. Stochastic Bayesian methods attempt to calculate the posterior probability of process parameters based on a prior probability of parameters and a likelihood function. Although the use of importance sampling techniques helps to improve efficiency, these methods are still computationally expensive. Optimization algorithms try to find optimal values of parameters that minimize a cost function. Gradients of the cost function with respect to process parameters are practically approximated by finite differences of perturbed integrations, and the choice of the cost function is critical. Two advanced data assimilation method: the adjoint method and the ensemble Kalman Filter method, seem to be promising methods for parameter estimation. Further studies are still needed to apply these data assimilation methods to calibrate climate models. In this study, the adjoint method will be further applied to estimate process parameters on climate timescales. A detailed introduction of the adjoint method and synchronization technique which can be used to extend the assimilation window will follow in chapter 3.

# Chapter 3

## Methodology

This chapter intends to provide a theoretical basis for this study. Based on the Lorenz (1963) system, mathematical formulation of variational data assimilation and the adjoint method which is practically exploited to solve the variational data assimilation problem are introduced in detail. Then, the limitation of the adjoint method is presented, and several methods which can be used to extend the feasible assimilation window are reviewed. Here, the dynamical state and parameter estimation method will be exploited in this study due to its potential to overcome the limited assimilation window problem and its easy implementation.

### 3.1 Mathematical formulation of variational data assimilation

The mathematical formulation of variational data assimilation involves minimizing a cost function which measures the distance between the model simulation and the observations subjected to a set of constraints (govern equations, boundary condition, .etc.). In practice, background terms for the control variables, including initial conditions, boundary conditions and poorly known parameters, also appear in the cost function to deal with the under-determined problem. The background terms usually have quadric forms which derive from the Gaussian assumption for the probability density function. The optimal values of the control variables can be found using an unconstrained minimization algorithm (e.g., Conjugate Gradient method, Quasi-Newton method) providing the information of the cost function and its gradient with respect to the control variables.

The Lorenz model (1963) provides a practical test case for data assimilation and is widely used in predictability and data assimilation studies (Gauthier, 1992; Miller et al., 1994; Pires et al., 1996; Stensrud and Bao, 1992). It shares some fundamental properties of the atmosphere such as the occurrence of regime behavior. Here, the Lorenz model is used to formulate variational data assimilation and to present limitations of the adjoint method. This system consists of three ordinary differential equations:

$$\dot{x} = \sigma(y - x) \quad (3.1)$$

$$\dot{y} = \rho x - y - xz \quad (3.2)$$

$$\dot{z} = xy - \beta z \quad (3.3)$$

where  $\sigma$  is the Prandtl number,  $\rho$  is the Rayleigh number and  $\beta$  is the domain aspect ratio. Given an initial condition  $(x_{\text{ini}}, y_{\text{ini}}, z_{\text{ini}}) = (x_0, y_0, z_0)$  and the parameters  $(\sigma, \rho, \beta)$ , the system will have a unique solution. If model variables are observed in time  $t$ :

$$(x(t), y(t), z(t)) = (x_{\text{obs}}, y_{\text{obs}}, z_{\text{obs}}) \quad (3.4)$$

and we may expect that the observations don't match the model solution. The system becomes over-determined. In practice, we cannot provide an accurate initial condition and the observations may also contain errors. For the purpose of parameter estimation, the parameter

$\rho$  is also assumed to have error. Therefore, the model simulation is different with the available observations. The govern equations (3.1)-(3.3) with a biased parameter  $\rho$ , an initial condition with errors, and the observations with small errors are rewritten as following:

$$\dot{x} = \sigma(y - x) \quad (3.5)$$

$$\dot{y} = (\rho + \delta\rho)x - y - xz \quad (3.6)$$

$$\dot{z} = xy - \beta z \quad (3.7)$$

$$X_{ini} = (x_{ini}, y_{ini}, z_{ini}) = (x_0 + \delta x, y_0 + \delta y, z_0 + \delta z) \quad (3.8)$$

$$X_{obs} = (x(t), y(t), z(t)) = (x_{obs} + \varepsilon_x, y_{obs} + \varepsilon_y, z_{obs} + \varepsilon_z) \quad (3.9)$$

Usually, the observations have much smaller errors than the model simulation. The problem now is to find optimal values of initial condition  $X_{ini}$  and parameter  $\rho$  which make model simulation as close as possible to the observations. This goal is achieved by minimizing the following quadratic cost function:

$$J(X_{ini}, \rho) = \frac{1}{2} \sum_{t=1}^n \frac{(M(X_{ini}, \rho) - X_{obs})^2}{\varepsilon^2} + \frac{1}{2} \frac{(X_{ini} - X_b)^2}{\delta^2} + \frac{1}{2} \frac{(\rho - \rho_b)^2}{\sigma_\rho^2} \quad (3.10)$$

where  $M$  represents the model operator. The symbols:  $\varepsilon$ ,  $\delta$  and  $\sigma_\rho$  represent uncertainties of the observations, the initial condition and parameter  $\rho$ .  $X_b$  and  $\rho_b$  are background values for the initial condition and the parameter  $\rho$ . In this formulation, the model is assumed to be perfect, and the variational data assimilation problem is called a strong-constrained problem in which the final solution will exactly satisfy the governing equations (3.5)-(3.7). In equation (3.10), the cost function uses a quadric form which is consistent with the Gaussian assumption from Bayes' theorem as equation (2.5). The first term on the right-hand-side of the equation (3.10) measures the model-observations misfits and the second and third terms are background terms depending on the relative accuracy of the initial condition and parameter  $\rho$ . For some implementations, other penalty terms like gravity wave penalty terms (Zou et al., 1993), nudging coefficient terms (Abarbanel et al., 2010), may also be added to the cost function to reduce their values during the optimization.

### 3.2 The adjoint method

Although several methods exist for minimizing the cost function (3.10) such as genetic algorithms (Evensen, 2009; Jackson et al., 2004) and penalty algorithms (Le-Dimet and Talagrand, 1986), a practical way is the adjoint method due to the large dimension of control variables in atmospheric models and oceanic models. The adjoint model is a powerful tool for directly and efficiently computing sensitivities of model outputs with respect to model inputs independent of variable dimensions. A detailed introduction of the adjoint method and its applications can be found in Errico (1997).

#### 3.2.1 The tangent linear model and its adjoint

In a nonlinear system, usually, the adjoint model is the adjoint of the tangent linear model (TLM). Figure 3.1 gives a schematic diagram of the tangent linear model. The black line represents a trajectory of  $X^r(t)$  starting from an initial condition  $X_0$ . The blue line  $X^p(t)$  is a

perturbed trajectory with a small perturbation  $\delta X$  on the initial condition  $X_0$ . After integrating for a period  $t$ , the two trajectories will be different, and the difference  $\Delta X$  can be calculated by finite differences of these two integrations:

$$\Delta X = X^p(t) - X^r(t) = M_{0 \rightarrow t}(X_0 + \delta X) - M_{0 \rightarrow t}(X_0) \quad (3.11)$$

Here,  $M_{0 \rightarrow t}$  represents a forecast model which evolves the model state  $X$  to specific time  $t$ . If the system is linear or for a nonlinear system linear dynamic dominates the integration, the tangent linear approximation based on the reference trajectory  $X^r(t)$  can also be used to represent the evolution of the perturbation  $\delta X$ . This correspondence can be derived with first-order Taylor series:

$$\Delta X = M_{0 \rightarrow t}(X_0 + \delta X) - M_{0 \rightarrow t}(X_0) \approx \frac{\partial(M_{0 \rightarrow t}(X_0))}{\partial X_0} \times \delta X = \mathbf{M}_{0 \rightarrow t}(X_0) \times \delta X \quad (3.12)$$

where  $\mathbf{M}_{0 \rightarrow t}(X_0)$  is the so-called tangent linear model depending on the reference trajectory  $X^r(t)$ . It should be noted that the tangent linear model  $\mathbf{M}_{0 \rightarrow t}(X_0)$  is a perturbation forecast model and it is exploited to predict the evolution of a small error  $\delta X$ , rather than the model state  $X$ . Although the tangent linear model is time-dependent through its dependent on the time evolution of the reference trajectory  $X^r(t)$ , it is a linear model because the tangent linear model in equation (3.12) doesn't depend on the prognostic variable  $\delta X$ .

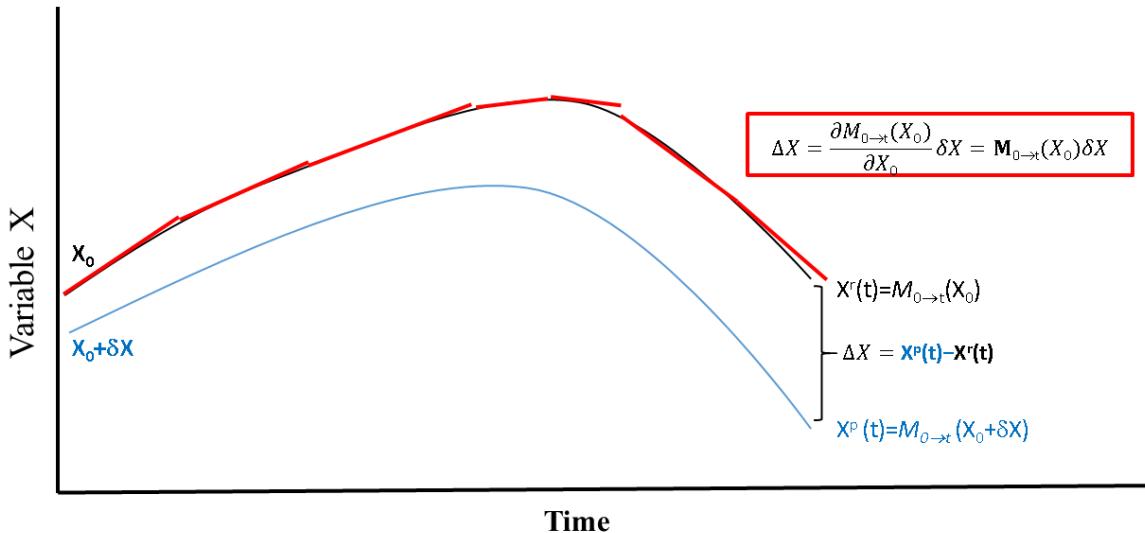


Figure 3.1. A schematic diagram of the tangent linear model. The black line represents a reference trajectory  $X^r(t)$ . The blue line represents a perturbed trajectory  $X^p(t)$ . The red stair-like line represents a tangent linear approximation based on the reference trajectory. The symbols  $M_{0 \rightarrow t}(X_0)$  and  $\mathbf{M}_{0 \rightarrow t}(X_0)$  represent a model forecast operator and the tangent linear model, respectively.

In practice, it is much more useful to derive sensitivity of a specific target ( $\Delta X$  in this case) with respect to the initial perturbation ( $\delta X$ ) or the model parameters, the reverse process of what the tangent linear model does. This can be derived by transpose of equation (3.12):

$$\delta X = \mathbf{M}_{t \rightarrow 0}^T(X_0) \times \Delta X \quad (3.13)$$

and  $\mathbf{M}_{t \rightarrow 0}^T(X_0)$  is the adjoint of the tangent linear model which performs as a backward integration from  $t$  to 0.

As equation (3.13) shows, the benefit of the adjoint method is that the computational cost is independent of the number of the control variables  $\delta X$ . Given a target  $\Delta X$  depending on the model state, sensitivities with respect to all the control variables  $\delta X$  can be directly derived by integrating the adjoint model  $\mathbf{M}_{t \rightarrow 0}^T(X_0)$  backward once.

For the Lorenz system, the equations of tangent linear model with respect to a reference trajectory  $X_r(t) = (x(t), y(t), z(t))$  are given as:

$$\begin{pmatrix} \delta \dot{x} \\ \delta \dot{y} \\ \delta \dot{z} \end{pmatrix} = \begin{pmatrix} -\sigma & \sigma & 0 & 0 \\ \rho - z(t) & -1 & -x(t) & x(t) \\ x(t) & y(t) & -b & 0 \end{pmatrix} \begin{pmatrix} \delta x \\ \delta y \\ \delta z \\ \delta \rho \end{pmatrix} \quad (3.14)$$

and the adjoint model of this system is:

$$\begin{pmatrix} \delta \dot{x}_{ad} \\ \delta \dot{y}_{ad} \\ \delta \dot{z}_{ad} \\ \delta \rho \end{pmatrix} = \begin{pmatrix} -\sigma & \rho - z(t) & x(t) \\ \sigma & -1 & y(t) \\ 0 & -x(t) & -b \\ 0 & x(t) & 0 \end{pmatrix} \begin{pmatrix} \delta x_{ad} \\ \delta y_{ad} \\ \delta z_{ad} \\ \delta \rho \end{pmatrix} \quad (3.15)$$

To minimize the cost function (3.10), the adjoint model is used to compute sensitivities of the cost function with respect to the initial state  $(x_{ini}, y_{ini}, z_{ini})$  and parameter  $\rho$ . With the adjoint model, gradients of the cost function (3.10) can be computed as:

$$\frac{\partial J(X_{ini}, \rho)}{\partial (X_{ini}, \rho)} = \sum_{t=n}^1 \mathbf{M}^T \left( \frac{M(X_{ini}, \rho) - X_{obs}}{\delta^2} \right) + \frac{x_{ini} - x_b}{\sigma^2} + \frac{\rho - \rho_b}{\sigma_\rho^2} \quad (3.16)$$

where  $\mathbf{M}^T$  is the adjoint operator. The difference between the model simulation and the observations is used as  $(\delta x_{ad}, \delta y_{ad}, \delta z_{ad})$  on the right-hand side of equation (3.15) to drive a backward integration of the adjoint model and the gradients of the cost function  $J$  with respects to the parameter  $\rho$  and the initial conditions  $X_{ini}$  can be computed.

### 3.2.2 The minimization algorithm

One of the important parts of the adjoint method consists of minimizing the cost function  $J$  by an iteratively unconstrained minimization algorithm using the gradient information provided by the adjoint model. In this study, a Quasi-Newton method based on Fletcher and Powell (1963) is exploited. The basic procedure of this method for minimizing the cost function  $J(C)$  is described below:

- 1) Starts with an initial guess of the control vector  $C_0$  (including the initial state  $X_{ini}$  and parameter  $\rho$  in the Lorenz system) and initialize the inverse Hessian matrix  $H_0$  with an identity matrix.
- 2) Set  $k=0$ , integrate the forward model and the adjoint model backward to compute the gradient of the cost function  $g_0$  based on equation (3.16) and set the searching direction:

$$d_0 = -H_0 g_0 \quad (3.17)$$

- 3) For  $k=1,2,\dots,n$ , iteratively search the minimum of the cost function  $J(C)$ . This is done through a line search of a proper positive scalar  $\alpha_k$  which minimizes  $J(C_k + \alpha_k d_k)$ . A new control vector  $C_{k+1}$  is obtained as:

$$C_{k+1} = C_k + \alpha_k d_k \quad (3.18)$$

- 4) Integrate the adjoint model backward with the new control vector  $C_{k+1}$  and compute a new gradient  $g_{k+1}$  based on equation (3.16).

$$g_{k+1} = \nabla J(C_{k+1}) \quad (3.19)$$

The convergence criteria is checked here. If  $\|g_{k+1}\| \leq \varepsilon$ , where  $\varepsilon$  is a small number, the optimization will be terminated and  $C_{k+1}$  is taken as the optimal solution. Otherwise, the process continues.

- 5) Correct the approximation of the inverse of the Hessian matrix  $H_{k+1}$  and compute a search direction  $d_{k+1}$  for the next iteration:

$$H_{k+1} = H_k + \frac{p_k p_k^T}{p_k^T q_k} - \frac{H_k q_k q_k^T H_k}{q_k^T H_k q_k} \quad (3.20)$$

$$p_k = C_{k+1} - C_k \quad (3.21)$$

$$q_k = g_{k+1} - g_k \quad (3.22)$$

$$d_{k+1} = -H_{k+1} g_{k+1} \quad (3.23)$$

And setting  $k=k+1$ , this process is repeated from step 3).

### 3.2.3 The optimization process

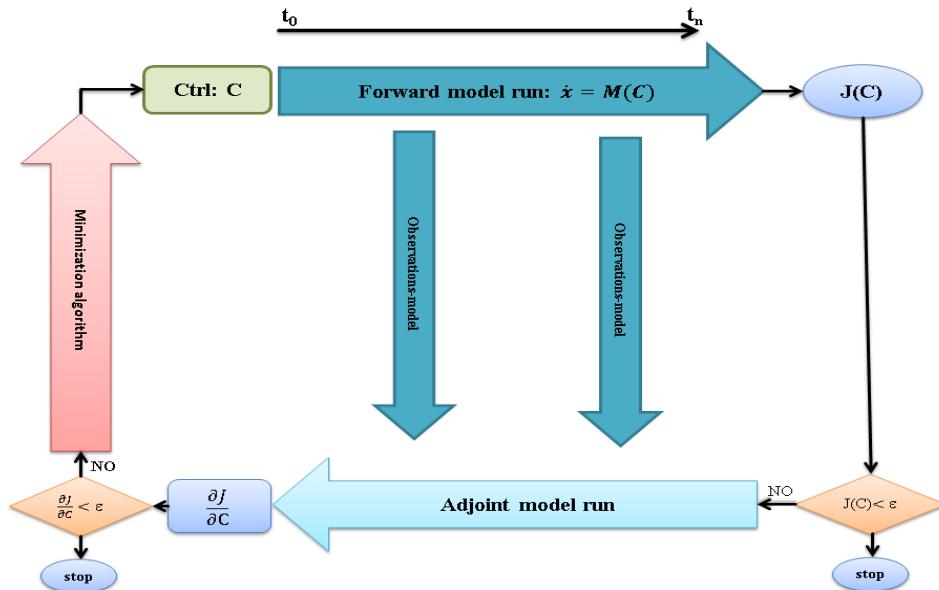


Figure 3.2. A schematic diagram of the recursive minimization process in the adjoint method.

With the gradients provided by the adjoint model, a minimization algorithm is adopted to iteratively reduce the cost function  $J(C)$ . Figure 3.2 gives a schematic diagram of the recursive minimization process in the adjoint method. Three stages involve in each iteration: 1) with a first guess of control vector  $C_0$  including the initial condition  $X_{ini}$  and parameter  $\rho$ , the model is integrated forward from  $t_0$  to  $t_n$ . The model-observation misfits are computed and stored

when the observations are available and the cost function  $J(C)$  is evaluated after the forward integration; 2) integrate the adjoint model from  $t_n$  to  $t_0$  with external forcing terms depending on the model-observation misfits. The gradients of the cost function  $J(C)$  with respect to the control vector  $C$  are computed; 3) providing the gradients and the cost function, a gradient descent algorithm is exploited to reduce the cost function  $J(C)$  and a new control vector  $C$  is provided for the next iteration. This recursive data assimilation process will be exit when the cost function  $J(C)$  or the gradients  $\frac{\partial J}{\partial C}$  is considerable small. At that time, the adjoint method successfully finds a model solution that minimizes the distance between the model simulation and the observations.

In equations (3.14) and (3.15), the tangent linear model and its adjoint model are presented in continuous forms. However, this is not relevant to real geophysical models in which the adjoint of the discretized model is usually used rather than the discretization of a continuous adjoint model such as equation (3.15). Because in an atmosphere or ocean general circulation model, some of the parameterizations such as vertical mixing usually have very complex forms and it is impossible to derive continuous adjoint equations. Practically, the adjoint of the discretized model is used, and some automatic differentiation tools such as TAF have been developed to automatically generate the tangent linear model and the adjoint model based on the discretized models.

In this study, the adjoint of the discretized model is used. Equations (3.1)-(3.3) are integrated with a fourth-order Runge-Kutta scheme and a time-step of  $\Delta t=0.01$ . The tangent linear model and its adjoint are generated by TAF. The three parameters are set to classical values:  $\sigma=10.0, \rho=28.0, \beta=8/3$ , which produce chaotic behavior (Nese et al., 1987) with Lyapunov exponents of (0.93, 0.00, -14.60). The model is spun up for  $10^4$  time units (TUs) with the final state (12.45260, 13.16454, 31.38284) used as an initial state for all assimilating experiments below.

The cost function used for the Lorenz system is as follow:

$$J(x_0, y_0, z_0, \rho) = \frac{1}{N} \left[ \sum_{m=1}^N (x(m) - x^{obs}(m))^2 + \sum_{m=1}^N (y(m) - y^{obs}(m))^2 + \sum_{m=1}^N (z(m) - z^{obs}(m))^2 \right] \quad (3.17)$$

in which the control variables are the initial state  $(x_0, y_0, z_0)$  and the Rayleigh number  $\rho$ .

### 3.3 Limitations of the adjoint method

The adjoint model is based on the tangent linear approximation, and the usefulness of the adjoint model is limited by the nonlinear system. The stronger the nonlinearity and the longer the integration time, the less accurate the tangent linear approximation will become. Therefore, the feasible assimilation window of the adjoint method is limited in a nonlinear system. Based on the Lorenz system, this limitation is presented from the cost function and the adjoint sensitivity.

### 3.3.1 Behaviour of the cost function

Previous studies (Gauthier, 1992; Miller et al., 1994; Pires et al., 1996; Stensrud and Bao, 1992) showed that the behavior of the cost function is strongly dependent on integration time. Pires et al. (1996) gave an estimation of the feasible assimilation window depending on first guess errors and the leading Lyapunov exponent  $\lambda$  of the nonlinear system:  $\tau \equiv -\ln(e)/(2\lambda) \leq 2$  TUs, where  $e$  is first guessed errors along the most unstable direction. Figure 3.3 displays the cost function for an integration of 2 TUs (dashed black line) and 5 TUs (solid black line) depending on the parameter  $\rho$ . For 2 TUs, the cost function is smooth, and the minimization will iteratively find the global minimum. While for 5TUs, dozens of secondary minima appear, and the gradient decent minimization algorithm is likely trapped into one of the secondary minima. The red squares and circles in Figure 3.3 represent two cases in which the optimization luckily and successfully reach the global minimum while the black markers represent three cases in which the optimization is trapped into local minima. With the extending assimilation window, more secondary minima will occur and the cost function will look like white noise (see Figure 6.1 in (Evensen, 2009)). This is caused by the chaotic behavior of the nonlinear system. Every secondary minimum represents a separation of two nearby trajectories and may close again after a period (Pires et al., 1996) for the Lorenz system. As Figure 3.5 shows, two nearby trajectories first separate at 15 TUs and go close again between 16-18 TUs.

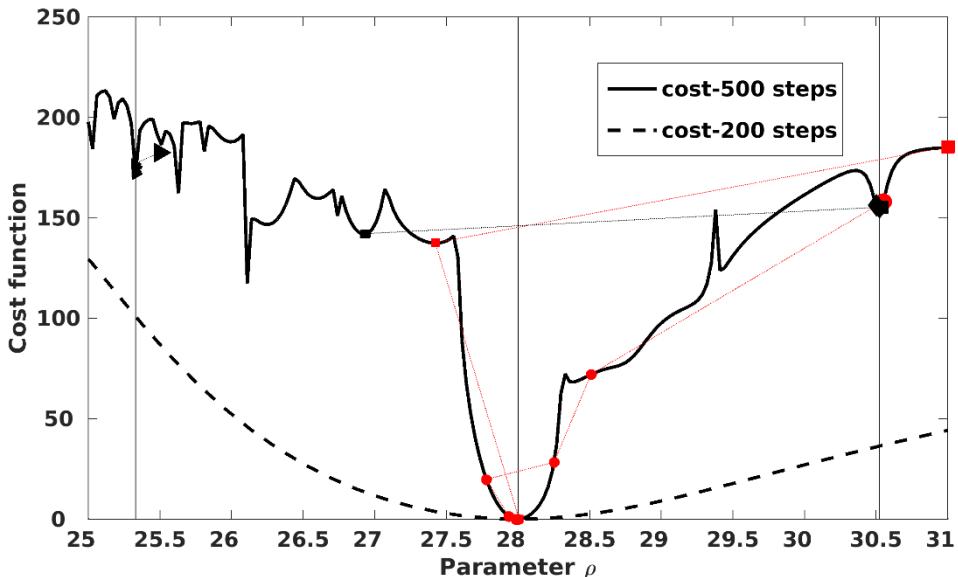


Figure 3.3. The cost function for an integration time of 5 TUs (black solid line) and 2 TUs (black dashed line). The red markers represent cases which reach the global minimum after the optimization and the black markers represent cases which are trapped into one of the secondary minima. The vertical lines indicate the minima after the optimization.

### 3.3.2 Behaviour of the adjoint sensitivities

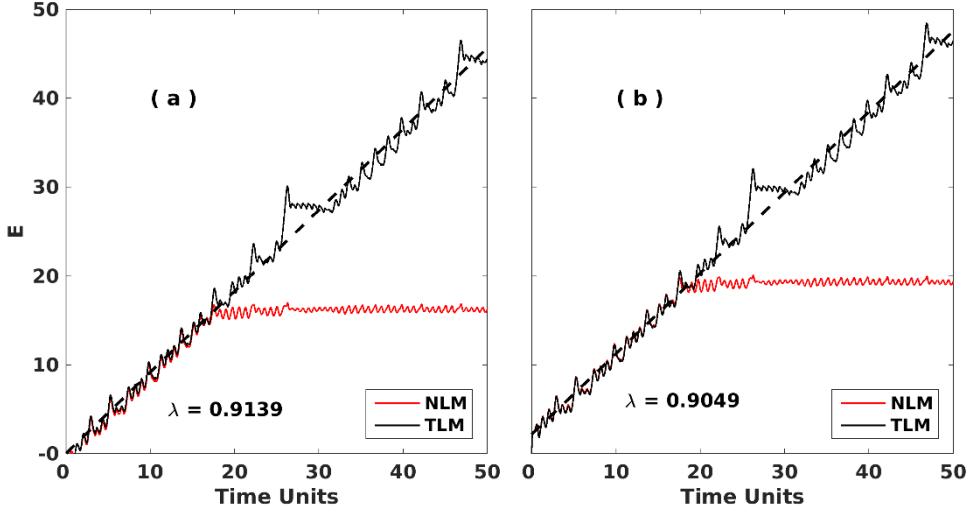


Figure 3.4. (a) The instantaneous error growth with small perturbations on parameter  $\rho$ : the red line is computed based on a reference integration and a perturbed integration of the nonlinear model; the black line is from the tangent linear model; (b) is similar to (a) but the perturbations are on the initial state. The black dashed line is a linear fit to the tangent linear error growth. The perturbation size is  $10^{-5}$ .

From the point view of predictability, the chaotic behavior of the nonlinear system acts as: two nearby trajectories will gradually separate starting from the most unstable modes with time integration and the average separating rate can be represented by leading Lyapunov exponent (Nese et al., 1987). When separation occurs, the prediction is not controlled by the initial state and the prediction limit is reached. The instantaneous local error growth caused by the initial errors (Rui-Qiang et al., 2008) and the parameter errors defined as equation (3.18) and the error growth from the tangent linear model is used to illustrate the limitation of the adjoint model. The bracket in equation (3.18) represents ensemble mean. For the nonlinear error growth rate, we use  $10^6$  ensemble members, and the perturbations are assumed to be random with an amplitude of  $\delta=10^{-5}$  on the initial state and parameter  $\rho$ .

$$E(x(t_0), \delta(t_0), \tau) = \left[ \ln \left( \frac{\|J(t_0+\tau)\|}{\|J(t_0)\|} \right) \right] \quad (3.18)$$

The red line and black line in Figure 3.4 (a) represent the instantaneous local error growth of the nonlinear model and the tangent linear model with small perturbations on parameter  $\rho$ . Linear fitting is overlaid as the black dashed line. The local errors grow exponentially with a rate around 0.9139 for both the nonlinear model and the tangent linear model for the initial 15 TUs. The local error growth will finally saturate when the trajectories separate due to the nonlinearity (see Figure 3.5) while the tangent linear error keeps growing. The instantaneous local error growth for the parameter  $\rho$  error is similar to that of the initial perturbation due to the chaotic nature of the nonlinear system. It should be noted that the perturbation of parameter  $\rho$  used here doesn't significantly change the leading Lyapunov exponent and therefore the chaotic nature of the nonlinear system. When separation occurs,

the tangent linear approximation breaks and the sensitivities compute by the adjoint model will only reflect the chaotic nature of the system and are different to that derived from finite differences of perturbed model integrations (Köhl and Willebrand, 2002; Lea et al., 2000). The separation time is directly related to perturbation sizes and the local leading Lyapunov exponent of the chaotic system. For real data assimilation problems, the error of the first guessed value is considerably large, the model simulation and the observations may separate within a short period which prevents the optimization converge to the global minimum.

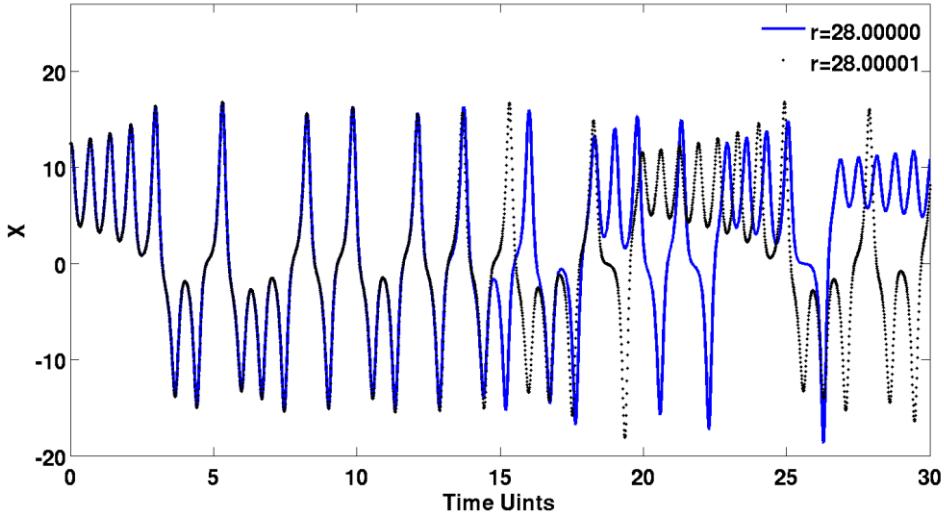


Figure 3.5. Two trajectories starting with the same initial state but slight different values of parameter  $\rho$ :  $\rho=28.0$  (the blue line) and  $\rho=28.00001$ (the black dotted line).

As analyzed above, the limitation of the adjoint method is that: with extending assimilation window, multiple minima occur in the cost function; the adjoint sensitivity will show an exponentially increase which cannot lead the optimization to the global minimum and the gradient-based minimization algorithms are likely to be trapped into secondary minima. Therefore, the feasible assimilation window is limited by the predictability of the nonlinear system. Although there are cases in which the optimization may reach the global minimum (see Figure 3.3), these are only lucky cases. Further extending the assimilation window will lead to more secondary minima and greatly increase the possibility of being trapped into a local minimum. When the assimilation is quiet long such as for climatological timescale of atmosphere models, gradients computed by the adjoint model will increase exponentially to infinite and are useless for the optimization. Regularizations are needed to extend the feasible assimilation window.

### 3.4 Extending the feasible assimilation window

Among the several methods trying to extend the feasible assimilation window are the quasi-static variational assimilation method (Pires et al., 1996), the statistical variational assimilation method (Köhl and Willebrand, 2002; Sugiura et al., 2008) and the dynamical state and parameter estimation method (Abarbanel et al., 2010; Abarbanel et al., 2009). The former one tries to avoid the secondary minima by tracking the absolute minimum with a progressively increasing assimilation window. Due to the chaotic behavior of the nonlinear

model, the assimilation window will still be limited to a certain period. The statistical variational assimilation uses a separate adjoint model which describes the sensitivity of the mean state (Köhl and Willebrand, 2002) or long timescale processes (Sugiura et al., 2008) rather than a special realization of trajectory to assimilate statistical or long timescale information. For this method, constructing the separate adjoint model needs some skills and usefulness of the approximate adjoint model cannot be easily validated. The DSPE method avoids the occurring of secondary minimum by synchronizing chaos of the nonlinear system and the observation system by coupling the two systems with nudging terms. (Abarbanel et al., 2010; Abarbanel et al., 2009) advocate that the nudging term can be penalized and pushed to zero after the optimization recovering the original physics of the nonlinear system. This method only needs small changes to the traditional variational data assimilation, and the feasible assimilation window can be extended arbitrarily. Moreover, the adjoint model generated by TAF can be directly used.

### 3.4.1 Regularization with synchronization

The DSPE method extends the feasible assimilation window by synchronizing the chaotic behavior of the nonlinear system and the observations. Synchronization of chaos refers to a process in which two similar dynamical system, driven by the same inputs, produce the same outputs after a long time integration (Alvarez, 1996; Boccaletti et al., 2002) and has interesting applications in synchronizing chaotic circuits and communications (Boccaletti et al., 2002). In practice, synchronization is exploited to force one chaotic system to follow the same path of another chaotic system. Figure 3.6 gives a schematic diagram of identical synchronization for two dynamical systems. System 1 and system 2 represent two systems with the same govern equations but start with different initial states.  $X(t)$  and  $Y(t)$  from system 1 are used to drive system 2. After a long period, the two systems will have the same model state and the two systems are completely synchronized. System 1 is called a master system which corresponds to the observations system in this study and system 2 is a slave system.

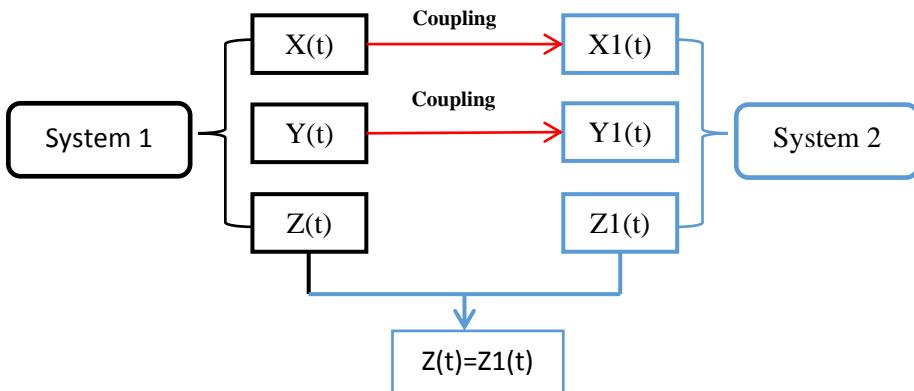


Figure 3.6. A schematic diagram of synchronizing two chaotic systems. System 1 and system 2 are coupled through  $(X_1, Y_1)$  and  $(X, Y)$ .

By considering data assimilation problems as a question of synchronization in which the nonlinear numerical model is synchronized to the observations, Abarbanel et al. (2010)

purposed the DSPE method. By synchronizing the nonlinear model and the observations, this method considerably extends the feasible assimilation window, which helps to estimate uncertain parameters and unobserved parts of the model state by using observations distributed over long time periods. With the updated parameters and model state, predictions are made. In the atmospheric community, synchronization is usually achieved by simple nudging method (Hoke and Anthes, 1976) and Lunkeit (2001) showed that two identical global atmospheric circulation models with many degrees of freedom could be completely synchronized with simple nudging terms once the nudging strength is strong enough. Equations (3.19)-(3.21) give an example of synchronization when a simple nudging term is added to equation (3.1). Based on the idea of Abarbanel et al. (2010), a penalty term for the nudging coefficient is added to the cost function as the second term on the right-hand side of equation (3.22). The penalty term tries to drive the nudging coefficient  $\alpha$  towards 0 during the optimization.

$$\dot{x} = \sigma(y - x) + \alpha(x - x^{obs}) \quad (3.19)$$

$$\dot{y} = \rho x - y - xz \quad (3.20)$$

$$\dot{z} = xy - \beta z \quad (3.21)$$

$$J(x_0, y_0, z_0, \rho, \alpha) = \{J_e\} + \{J_p\} = \left\{ \frac{1}{N} \left[ \sum_{t=1}^N (x(t) - x^{obs}(t))^2 + \sum_{t=1}^N (y(t) - y^{obs}(t))^2 + \sum_{t=1}^N (z(t) - z^{obs}(t))^2 \right] \right\} + \{W_\alpha^2 \alpha^2\} \quad (3.22)$$

(Abarbanel et al., 2010; Abarbanel et al., 2009) suggested that the initial value of the nudging coefficient  $\alpha$  should be strong enough to push positive Lyapunov exponents to negative values. Ideally, the penalty term in equation (3.22) should gradually reduce the nudging coefficient  $\alpha$  while keeping it still strong enough to avoid separations in each iteration. At the end of the optimization, the nudging coefficient  $\alpha$  is close to 0 and recovers the original physics of the model. In the framework of variational data assimilation, the penalty term of the nudging coefficient  $\alpha$  usually appears as a background term with a background value of  $\alpha=0$  and  $W_\alpha$  represents its uncertainty which is investigated here. The weighting factor  $W_\alpha$  will strongly influence the convergence of the nudging coefficient  $\alpha$ , hence it will influence the successfulness of this method especially when the assimilation window is rather long (e.g., climatology timescale).

### 3.4.2 Effects of nudging

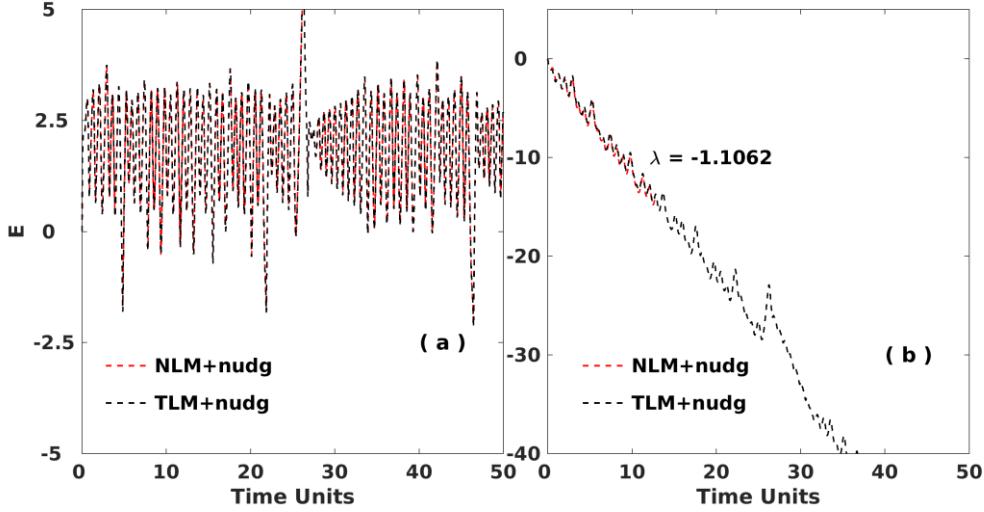


Figure 3.7. The instantaneous local error growth for perturbed initial state (b) and perturbed parameter  $\rho$  (a) with perturbation sizes of  $10^{-5}$  for the synchronized system (3.19)-(3.21): the black dashed lines are from the tangent linear model and the red dashed lines are from perturbed the nonlinear system integrations. The nudging coefficient is  $\alpha=20$ .

One of the important differences between parameters errors and initial state errors is that parameter errors have impacts on the model simulation at each time step while initial state errors affect model simulation indirectly through error propagation along with the model integration. The red dashed line and black dashed line in Figure 3.7 show the mean local error growth of the nonlinear model and the tangent linear model caused by initial state perturbations (b) and parameter  $\rho$  perturbations (a) for the slave system (3.19)-(3.21) with  $\alpha=20$ . As Figure 3.7(b) indicates, the nudging term has moved the leading local Lyapunov exponent from 0.9139 to -1.1062 and initial errors are exponentially damped out with a rate around -1.1062 while parameter  $\rho$  perturbation still acts as a constant force that pushes the trajectory away from the observations as the black dashed line and the red dashed line in Figure 3.7(a) show. Unless we replace the model simulation with the system of the observations, the slave system can never be completely synchronized, and parameter  $\rho$  perturbation information will always remain in the slave system (3.19)-(3.21).

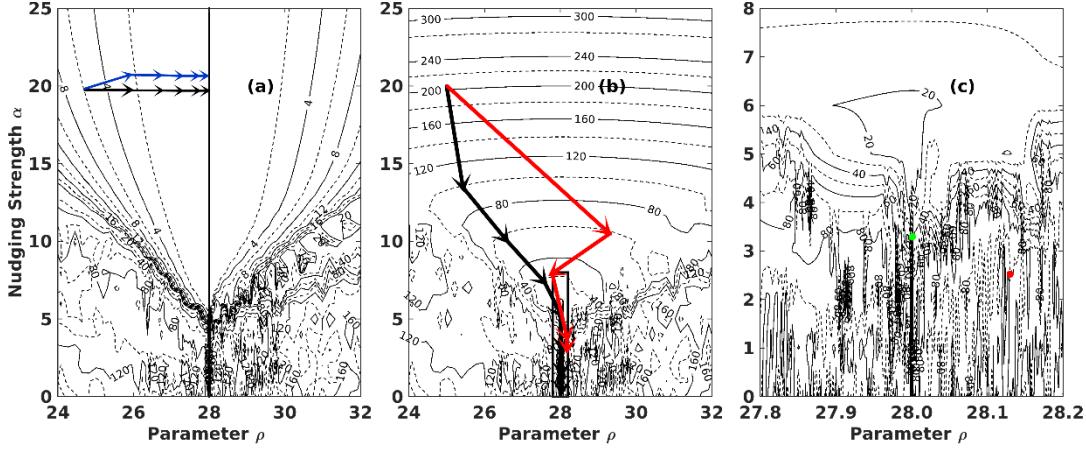


Figure 3.8. The cost function depending on nudging coefficient  $\alpha$  and parameter  $\rho$  without the penalty term (a) and with the penalty term and  $W_\alpha=1.0$  (b). The black rectangle area of (b) is zoomed in as (c). The estimated nudging coefficient  $\alpha$  and parameter  $\rho$  for expt1 and expt2 (in Chapter 4) are marked with the green and red filled circles. The arrows represent schematic optimization routes for different experiments. The integration time is 20 TUs. The cost functions are computed with error-free data.

Figure 3.8 shows the cost function depending on the nudging coefficient  $\alpha$  and the parameter  $\rho$  without the penalty term (a) and with the penalty and  $W_\alpha=1.0$  (b). When the nudging coefficient is larger than around 5, the cost function surface shows a regular core around zero parameter error without secondary minima that starts to widen with increasing coefficient for both cases. With increasing nudging coefficient, the cost function surface becomes regular for the given range of the parameter  $\rho$  (22.0-34.0) while the position of the global minimum for parameter  $\rho$  doesn't change as the thick black line in Figure 3.8 (a) shows. In Figure 3.8 (b), the inclusion of the penalty term reduces the global minimum from the thick black line in Figure 3.8 (a) to a single point (28.0, 0.0). Based on the idea of the DSPE method, the parameter should converge faster than the nudging strength goes down so that the nudging strength can avoid the occurrence of secondary minima. In Figure 3.8 (c), we zoom into the black rectangle region of Figure 3.8 (b) in which the nudging coefficient is smaller than 8, and the parameter  $\rho$  is around 28. The cost function is irregular even with very small parameter error when the nudging strength is smaller than 5. It is unlikely that the optimization can reach the global minimum (28.0, 0.0) and is therefore likely be trapped into secondary minima.

In the original DSPE method, the nudging strength should be strong enough to avoid the occurrence of secondary minima. The weighting factor  $W_\alpha$  which appears as the uncertainty of the nudging coefficient  $\alpha$  in variational data assimilation is likely to play an important role in tuning the nudging coefficient  $\alpha$  during the optimization and therefore significantly influences the success of the optimization. When adding the nudging term, the forward model has been altered. But the influence of parameter error remains as Figure 3.7 shows, and it affects the cost function as Figure 3.8(a) displays. For parameter estimation, we argue that constant nudging coefficients can be used without penalizing them. Providing enough observations, the

nudging term should at least be strong enough to push the positive Lyapunov exponents to negative values or stop the exponential error growth in the tangent linear model.

For the modified DSPE method, the slave system is still as equations (3.19)-(3.21) while the cost function is replaced with:

$$J(x_0, y_0, z_0, \rho) = \left\{ \frac{1}{N} \left[ \sum_{t=1}^N (x(t) - x^{obs}(t))^2 + \sum_{t=1}^N (y(t) - y^o(t))^2 + \sum_{t=1}^N (z(t) - z^o(t))^2 \right] \right\} \quad (3.23)$$

In this method, we only need to add nudging terms to the forward model with proper nudging coefficients. In the following chapter, this modified DSPE method will be investigated based on the Lorenz system.

# Chapter 4

## Assimilating experiments with the Lorenz system

To evaluate the performance of parameter estimation based on the classical DSPE method (Abarbanel et al., 2010; Abarbanel et al., 2009) and the modified DSPE method, we performed case studies with perfect observations and noisy observations. The influences of the nudging coefficient, sets of variables used for nudging and observation frequency, which impacts synchronization efficiency, are investigated. Throughout this study, we use the term “synchronization efficiency” to represent how well the nudging terms can recover the error-free signal from the noise and sparse observations.

The first guess of the control variables are  $(x,y,z,\rho)=(12.4473,11.2885,34.3449,24.5255)$  for all experiments. Unless declared, the assimilation window is 20TUs which is around ten times the feasible assimilation window (Pires et al., 1996). For noisy observations, the observation noise is assumed to have a uniform distribution with amplitudes: 1.37 for  $x$ , 1.56 for  $y$  and 4.35 for  $z$ , which corresponds to a signal-to-noise ratio of 20db. For the minimization, the maximum iteration number is set to 80 and the convergence criterion is a gradient norm below  $10^{-8}$ .

### 4.1 Case studies

#### 4.1.1 Perfect observations

In this section, identical-twin experiments with perfect observations are performed. The influence of the penalty term in the classic DSPE method (Abarbanel et al., 2010) is investigated through expt1, expt2 and expt3 which use a different weighting factor  $W_a$  (see table 4.1). In expt4, a constant nudging coefficient is used, and it is excluded from the control variables. Observations are provided every time step for all state variables. For perfect observations, the success of the optimization will be judged by the final parameter error and whether or not the convergence criterion is reached. The estimated values of the control variables and the nudging coefficient  $\alpha$  for all four experiments are listed in table 4.1.

Table 4.1. Parameter  $\rho$ , initial state and nudging coefficient  $\alpha$  after the optimization.

Expt No	$W_a$	Perfect observations					Noisy observations				
		$\alpha$	$\rho$	$x$	$y$	$z$	$\alpha$	$\rho$	$x$	$y$	$z$
Expt1	1	2.5189	28.1298	3.6793	15.9187	36.1799	0.70389	31.0058	12.3408	11.3836	34.3991
Expt2	0.01	3.2921	28.0008	12.8009	13.1395	31.1078	17.6873	27.9914	12.5594	12.8587	31.1276
Expt3	0.	20.848	27.9999	12.4526	13.1645	31.3828	98.1526	28.0148	14.1599	12.9026	30.8440
Expt4	---	20.000	27.9999	12.4526	13.1645	31.3828	20.0000	28.0423	12.4818	12.9669	31.1037

Figure 4.1 (a) and (b) shows the norm of  $\delta\rho$  and the nudging strength  $\alpha$  depending on iteration number. Both expt3 and expt4 have successfully retrieved parameter  $\rho$  within 30 iterations. The nudging coefficient  $\alpha$  increases slightly from 20 to 20.85 in expt3. For expt1 and expt2, the optimization is stopped by exceeding the maximum iteration number rather than satisfying the gradient norm criterion. The red and green circles in Figure 3.8 (c) show

the estimated parameter  $\rho$  and the nudging coefficient  $\alpha$  after the optimization for expt1 and expt2. Although the estimated values of parameter  $\rho$  are also close to 28 in these two experiments (see also table 4.1), chaos reemerges in expt1 and expt2 which can be inferred from the dramatical increase of gradient norm as Figure 4.1(d) displays. From table 4.1, it is noted that expt3 and expt4 also accurately estimate the initial state because observations are provided every time step while expt1 and expt2 fail. Even the maximum iteration number is set to 1000, expt1 and expt2 still cannot reach the global minimum (not shown here).

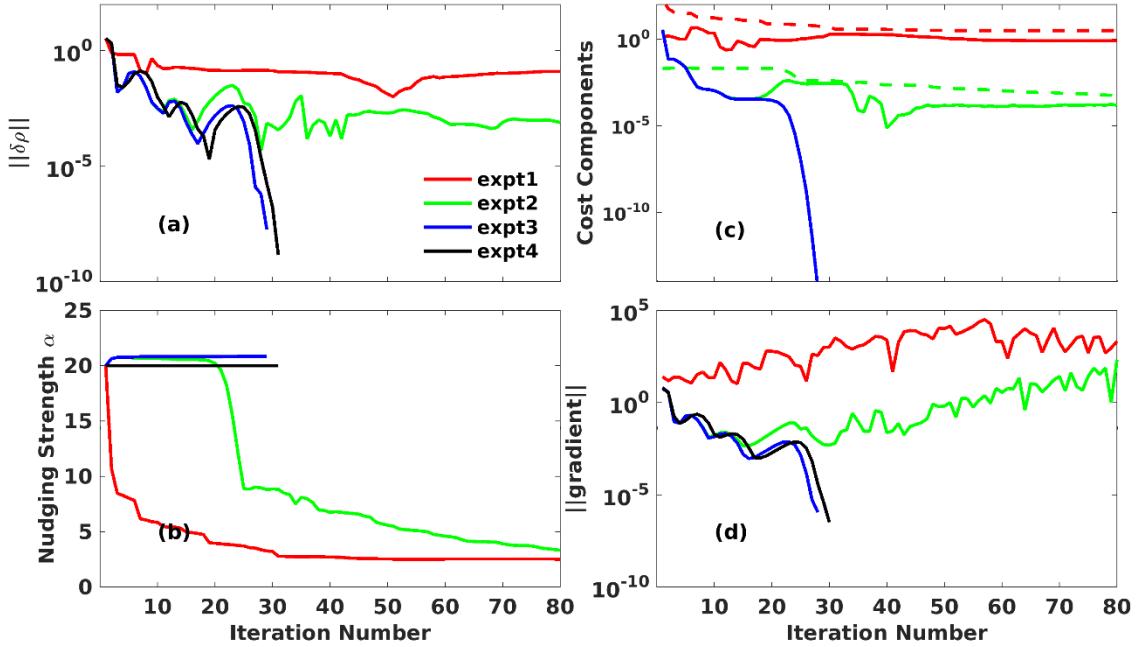


Figure 4.1. (a) norm of  $\delta\rho$ , (b) the nudging coefficient  $\alpha$ , (c) the cost function components  $J_p$  and  $J_e$ , and (d) norm of gradient, as function of iteration number for perfect observations. The line colors represent different experiments indicated by legend in (a). The dashed lines and solid lines in (c) represent  $J_p$  and  $J_e$  component of the cost function as in equation (3.22).

The different specification of  $W_a$  in expt1, expt2 and expt3 significantly influence the performance of the optimization. For simplicity, we use  $J_e$  to represent the model-observation misfit cost and  $J_p$  to represent the penalty term as the second term on the right-hand side of equation (3.22). Increasing the nudging coefficient  $\alpha$  decreases  $J_e$  and increases  $J_p$ . When minimizing the whole cost function, the optimization path is constrained by the balance between them. In Figure 4.1 (c),  $J_p$  is larger than  $J_e$  in expt1; the optimization reduces the nudging coefficient  $\alpha$  to reduce the whole cost function during the initial iterations. In expt2,  $J_e$  is comparable or larger than  $J_p$  in the first few iterations, and the optimization efficiently reduces  $J_e$  mainly by tuning the control variable  $\rho$  while only slightly increasing the nudging coefficient  $\alpha$ . After that,  $J_p$  dominates the cost function and the optimization reduces the cost function by quickly decreasing the nudging coefficient  $\alpha$  as Figure 4.1 (b) shows. When the nudging coefficient  $\alpha$  is decreased below a threshold value (around 10 in this case), a transition appears which is shown as a significant increase of the gradient norm in Figure 4.1 (d) starting from 3rd iteration in expt1 and the 22th iteration in expt2. The slave system

becomes chaotic again, and the estimated  $\rho$  shows no improvement after this transition. At this stage, the optimization reaches a balance between reducing the penalty term  $J_p$  and increasing the model-observation misfits  $J_e$ . The red arrows in Figure 3.8 (b) gives a schematic route for expt1 and expt2. In expt3, in which no penalty term is added for the nudging coefficient  $\alpha$ , the optimization slightly increases the nudging strength which implies that tuning the initial state and parameter  $\rho$  is the most efficient way to reduce the cost function. Optimization routes of expt3 and expt4 are provided as the blue arrows and black arrows in Figure 3.8 (a).

#### 4.1.2 Noisy observations

The four experiment setups are now tested with noisy observations. We use expt1n, expt2n, expt3n and expt4n to represent the new experiments. The optimization is successful if the gradient criterion is satisfied within the maximum iteration number.

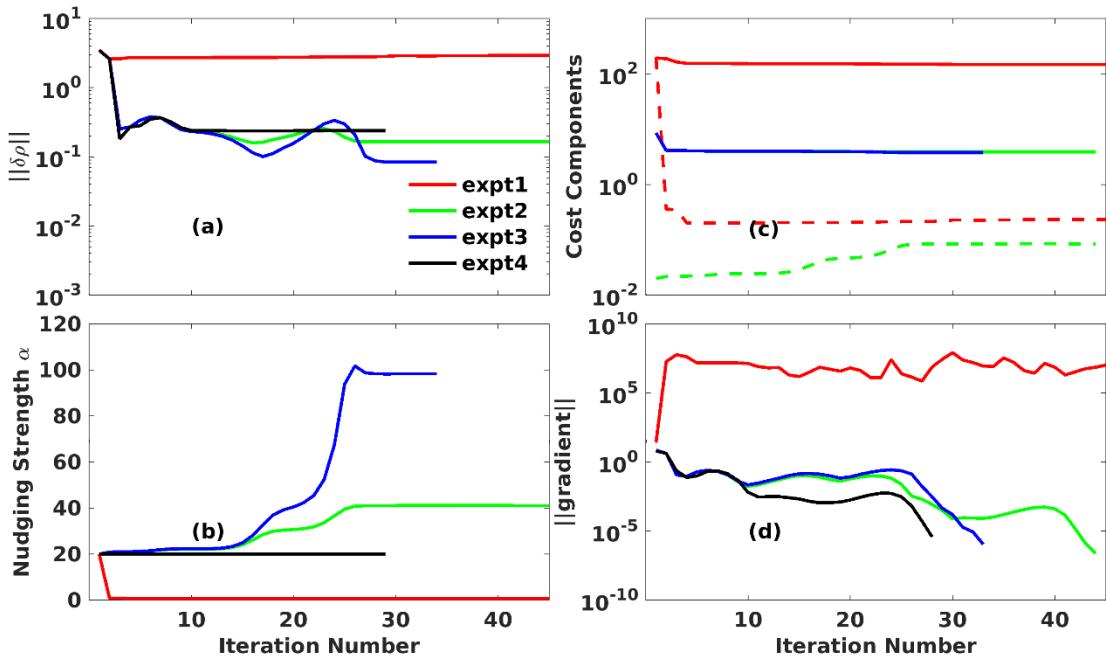


Figure 4.2. (a) norm of  $\delta\rho$ , (b) the nudging coefficient  $\alpha$ , (c) the cost function components  $J_p$  and  $J_e$ , and (d) norm of gradient, depending on iteration numbers. The line colors represent different experiments indicated by legend in (a). The dashed lines and solid lines in (c) represent  $J_p$  and  $J_e$  component of the cost function as in equation (3.22).

The norm of  $\delta\rho$  and the estimated values of the nudging coefficient  $\alpha$  are plotted in Figure 4.2 (a) and Figure 4.2 (b) and also listed in table 4.1. Only expt1n fails to retrieve parameter  $\rho$ . In expt2n and expt3n, the nudging coefficient  $\alpha$  is strongly increased while it is decreased in expt1n. Expt1n fails because the dramatic decrease of the nudging strength  $\alpha$  during the initial iterations, which makes the system chaotic again and multiple minima reoccur. Compared with Figure 4.1 (c), Figure 4.2 (c) shows that  $J_e$  stays much larger than  $J_p$  during the optimization in expt2n and expt3n. In this situation, the optimization will further

reduce the cost function by increasing the nudging coefficient  $\alpha$  to fit the noisy signal after the parameter  $\rho$  is accurately estimated. The gradient norm criterion is finally reached when the negative gradient of  $\alpha$  with respect to  $J_e$  and positive gradient of  $\alpha$  with respect to  $J_p$  cancel each other. At this stage, the effect of reducing  $J_e$  and increasing  $J_p$  reaches a balance. These results are similar to the findings by Bao (1993). However, since they investigated only assimilation windows much shorter than the limits of the validity of the linearized model and in our study the nudging coefficient  $\alpha$  is an essential factor that enables the feasibility of the optimization in a chaotic regime and therefore has additional constraints depending on the parameter error to keep the cost function regular as Figure 3.8 illustrates.

Although parameter  $\rho$  is successfully estimated in expt2n, expt3n and expt4n, expt2n and expt4n provide a better estimation of the initial state. In expt3n, the nudging strength is dramatically increased to 98.1526 which strongly erased the initial state errors, and the optimization provides a bad estimation of initial state with noisy observations.

The results of these case studies indicate that a successful optimization can be achieved only when the nudging strength is strong enough to avoid separation of the model simulation from the observed trajectory during the optimization. A proper  $W_a$  needs to be chosen in the classic DSPE method which is difficult to specify. Under the presents of noise in the observations, a large  $W_a$  or a no-existing penalty term leads to increasing of the nudging coefficient  $\alpha$  because the noise can only be fitted by the nudging terms. A small  $W_a$  leads to fast decay of the regularization, chaos reappears and stalls the optimization. Results of expt4 for both perfect observations and noisy observations show that the optimization works perfectly and efficiently with a constant nudging coefficient  $\alpha$ . In summary, we don't see a benefit of including the nudging coefficient as part of control variables if the goal is to estimate parameters in chaotic systems. The behavior of the slave system depends on the nudging coefficient  $\alpha$ , and the nudging coefficient  $\alpha$  is tuned by the weighting factor  $W_a$  for the penalty term. In the classical DSPE method, the weighting factor for the penalty term should be specified carefully which requires detailed knowledge of the cost function. While for the modified DSPE method, a minimum required nudging coefficient  $\alpha$  should be predefined which is relatively easy to determine.

With a constant nudging term, the initial state can also be reasonably estimated if accurate observations are provided in the time window before the initial information is erased. For long assimilation windows, the initial error will be erased for most of the assimilation window by the strong nudging terms and shows no significant influences on parameter estimation. In the following part, we will only provide results for the estimated parameter  $\rho$ .

## 4.2 Dependence on synchronization efficiency

With noisy observations, the nudging terms will introduce besides the true signal also noise into the system. Previously studies (Alvarez, 1996; Yang et al., 2006) have indicated that synchronization efficiency depends on the noise level, variables chosen for nudging, observational frequency and nudging strength. In this section, the influence of these factors on the parameter estimation will be investigated.

#### 4.2.1 Influence of observation noise

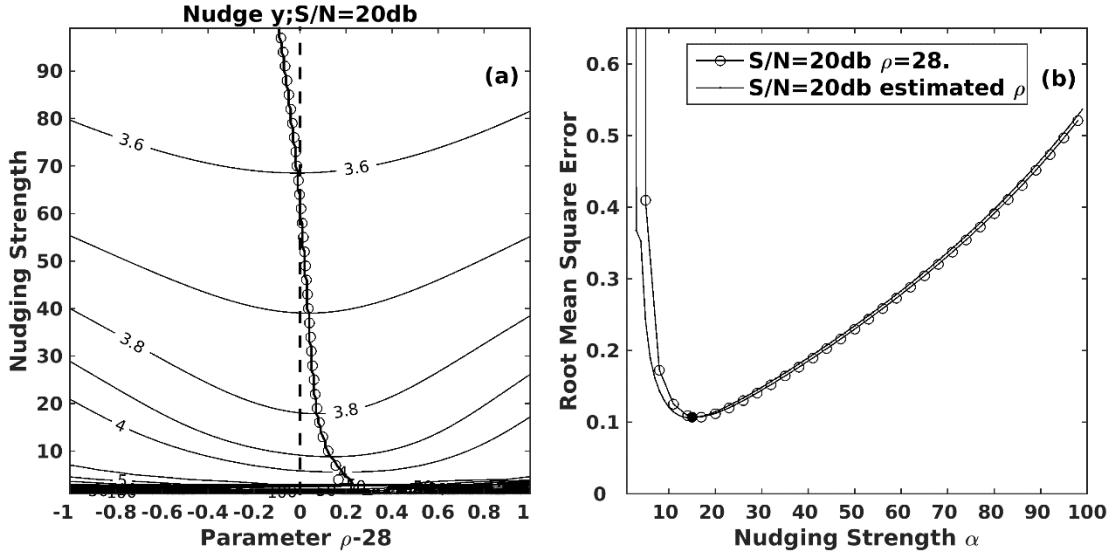


Figure 4.3. (a) The cost function depending on the parameter  $\rho$  and the nudging strength  $\alpha$  for the case of nudging  $y$ . The thick black lines represent global minima for a specific nudging strength  $\alpha$  and the black circles are the estimated parameter  $\rho$  from the optimization. (b) Root mean square errors compute from error-free observations with the slave system with true parameter (black circles) and optimized parameter (black line).

Equation (4.1) gives a general form of nudging with noisy observations and a parameter error in which observation noise and error-free signal are represented separately. With perfect observations, the nudging term will only introduce the true signal to the system and the difference between the slave system and observations will disappear when the parameter error becomes smaller. Parameters can be precisely estimated by variational data assimilation. With noisy observations, the nudging term will also introduce noise to the system as the third term on the right-hand side of equation (4.1) which degrades synchronization. The larger the nudging coefficient, the more noise will be introduced. There exists an optimal nudging strength which achieves the best ratio between true signal and introduced noise as shown in Figure 4.3 (b) and Figure 4.5. Here we investigate the influence of observation noise introduced by the nudging term on the parameter estimation. The variables used for nudging is  $y$ , and other sets of variables will be tested in next section. Experiments without penalizing nudging coefficient are repeated by gradually increasing nudging strength from 0 to 100. The assimilation window is 20 TUs.

$$\frac{\partial u}{\partial t} = f(u, \rho) + \alpha \times (u_{true} - u) + \alpha \times u_{noi} + \delta \rho \times \frac{\partial f(u, \rho)}{\partial \rho} \quad (4.1)$$

Figure 4.3(a) shows the cost function depending on nudging strength  $\alpha$  and parameter  $\rho$ . The global minima of the cost function for a specific nudging strength  $\alpha$  and the estimated parameter  $\rho$  are overlaid as the thick black line and black circles, respectively. When the nudging strength is larger than 5, the cost function surface is regular without any secondary minima. The Optimization always reaches the global minima by tuning parameter  $\rho$ . The

estimated parameter slightly depends on the nudging coefficient because the non-linearity of the system may lead to a rectifying effect of the noise introduced into the system by the nudging term or because parameter changes can accommodate for the observations noise, although the latter is less likely since the noise is symmetric. The RMSEs computed from error-free data with the slave system and the last iteration of every assimilation experiment are plotted in Figure 4.3(b). The two RMSEs curves are almost indistinguishable which indicates that the noise signal introduced by the nudging term could not be reduced further after the optimization. With the increasing nudging coefficient, the sensitivity of parameter is also reduced which can be seen from the decrease of the cost function range with the increasing nudging coefficient  $\alpha$  in Figure 4.3(a). To investigate the causes for residual error of the estimated parameter  $\rho$  under noisy conditions, three groups of experiments are performed. Expt5 uses the optimal nudging strength  $\alpha=15$  which leads to the best synchronization shown as the black filled circle in Figure 4.3(b), while in expt6  $\alpha=60$  is set which introduces two times more noise into the system. Expt 7 is similar to expt6 except that error-free  $y$  observations are used for nudging which will not introduce noise to the slave system. The same noisy observations are used to compute the cost function in these three experiments. We perform the three groups of experiments for assimilation windows from 2 TUs to 300 TUs.

As Figure 4.4(a) shows, when the assimilation window is smaller than 40TUs, the estimated parameter varies significantly because for a limited number of error realizations the mean over all errors will not be exactly zero even for symmetric noise and the estimated parameter can compensate for this change in mean. The shift of the mean gets smaller with the length of the assimilation window and beyond 40TUs, the estimated parameter is robust and the RMSEs in Figure 4.4(b) are close to their corresponding errors introduce by the nudging term in Figure 4.3(b). Expt 5 better retrieved the parameter than expt6, which indicates that the noise introduced by nudging term influences the accuracy of the estimated parameter. Expt7, which doesn't introduce noise signal from the nudging term best retrieved the parameter and also recovered the noise-free trajectory. After optimization, the whole nudging term is almost 0 in expt7 while it always exists in expt5 and expt6 due to the noisy observations. The results here show the noise introduced by the nudging term influences the accuracy of parameter estimation which cannot be reduced after optimization while a compensation of a shift in mean due to a limited number of error realizations can be largely reduced by extending the assimilation window. The results also indicate that the less the noise is introduced by the nudging terms, the more accurate the estimated parameter  $\rho$  will be.

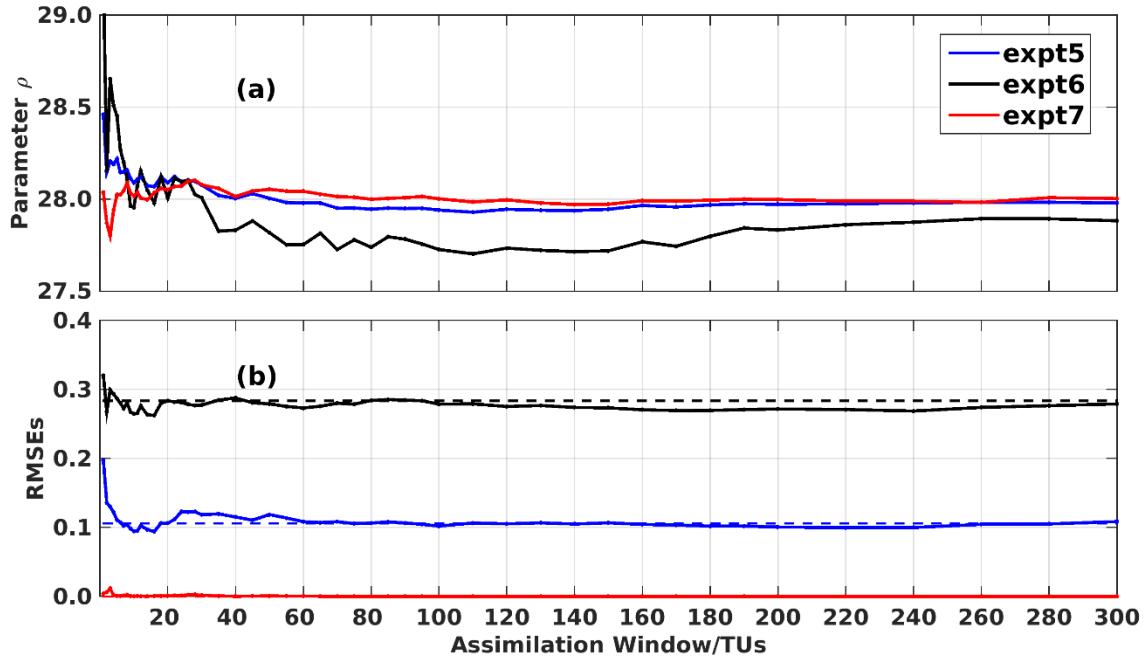


Figure 4.4. (a) The estimated parameter  $\rho$  depending on assimilation window for nudging  $y$ . Expt5 (blue line) uses a nudging coefficient  $\alpha=15$  with noisy  $y$  observations. Expt6 (black line) uses a nudging coefficient  $\alpha=60$  with noisy  $y$  observations and Expt7 (red line) uses a nudging strength  $\alpha=60$  with error-free  $y$  observations. (b) Root mean square errors for the above experiments. The dashed line represents the noise error introduced by nudging term shown in Figure 4.3(b).

#### 4.2.2 Experiments with different nudging variables

Figure 4.5 and Figure 4.3(b) show the RMSEs computed from error-free observations and the slave system synchronized by noisy observations for 200 TUs. The optimal nudging strength is marked with the filled markers. In agreement with Alvarez (1996), we found that the Lorenz (1963) system can be synchronized with six different sets of variables:  $x$ ,  $y$ ,  $xy$ ,  $xz$ ,  $xyz$ ,  $yz$ . Nudging  $z$  doesn't help to synchronize the system while it introduces noise to the system. Both nudging  $x$  and nudging  $y$  can synchronize the system and nudging  $y$  performs better than nudging  $x$ . When used together, it leads to the smallest RMSEs. Including nudge  $z$  leads to larger RMSEs than excluding nudging  $z$  when the nudging strength is strong enough for synchronization. Nudging  $xz$  leads to the largest RMSEs even with its optimal nudging strength. In this section, the influence of variables chosen for nudging on the parameter estimation is further investigated. This is done by using the setup of expt5 with the other five sets of nudging variables and gradually increasing the assimilation window from 2 TUs to 300 TUs. The optimal nudging strength is shown as the filled markers in Figure 4.5.

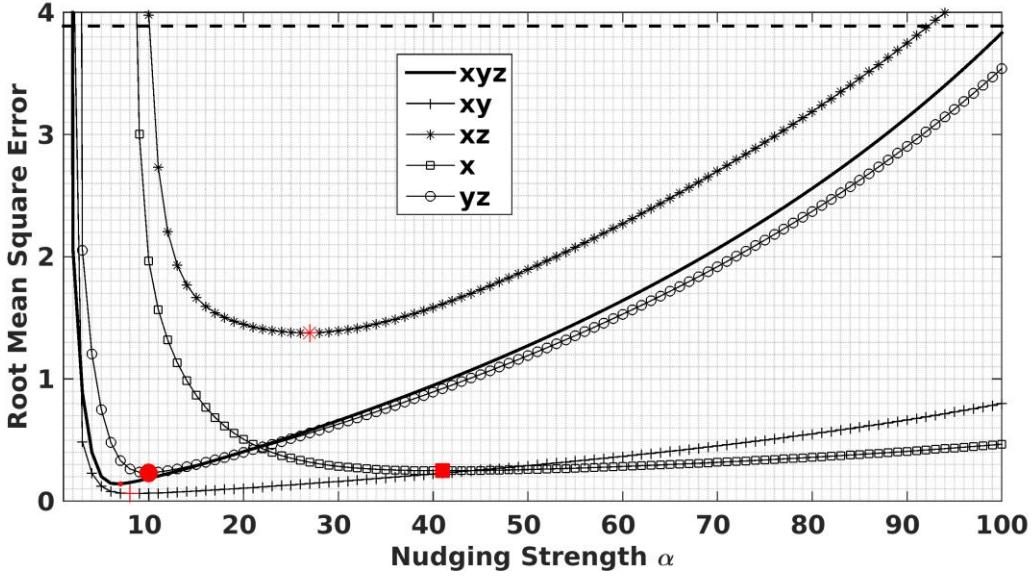


Figure 4.5. Root mean square errors computed from error-free observations and the slave system with different sets of variables chosen for nudging (see legend). The red markers represent the optimal nudging coefficient which best recovers the error-free signal from noisy observations. The dashed line is noise level.

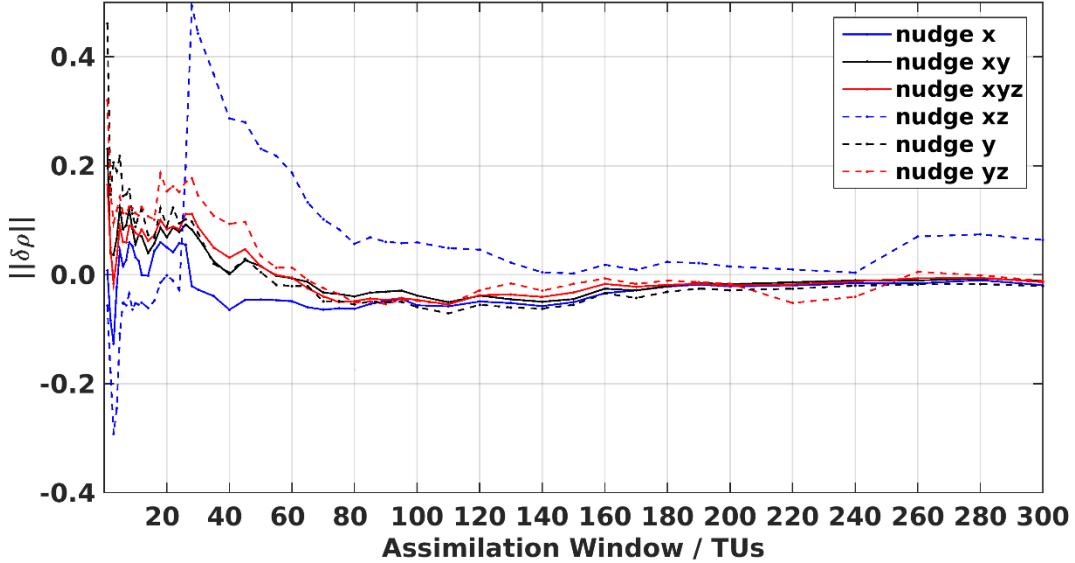


Figure 4.6. The estimated parameter  $\rho$  depending on the assimilation window with different sets of variables chosen for nudging (see legend).

Figure 4.6 shows the estimated parameter  $\rho$  for all sets of variables used for nudging depending on assimilation window. For windows smaller than 40TUs, the estimated parameter varies with the assimilation window and different nudging variables. When nudging x, xy, y and xyz, the estimated parameter  $\rho$  gradually converges and goes close to the true value with larger assimilation windows. Although the estimated parameter when nudging xz is also accurate for assimilation range from 140 TUs to 240 TUs, it is less accurate for other assimilation window ranges and it also shows some abrupt jumps around 25 TUs and

260 TUs because the noise introduced by nudging z degrades synchronization efficiency. Estimated parameter when nudging yz also slightly varies for the assimilation windows of 220 TUs and 240 TUs but less serious than for nudging xz.

Results here indicate that variables used for nudging also influence the accuracy of the parameter estimation. With different sets of variables chosen for nudging, the synchronization efficiency varies and hence the accuracy of the estimated parameter  $\rho$ . The different synchronization efficiency for different variables chosen for nudging reflects the relative importance of the variables for regularizing the unstable manifolds (Abarbanel et al., 2010; Yang et al., 2006). In real implementations, it is hard to determine the unstable manifolds due to the high complexity and large dimension of the atmospheric system. We suggest to nudging all variables which regularizes all manifolds and find out the optimal nudging strength to produce an effective synchronization, although nudging irrelevant variables may introduce unnecessary additional noise.

#### 4.2.3 Influence of observation frequency

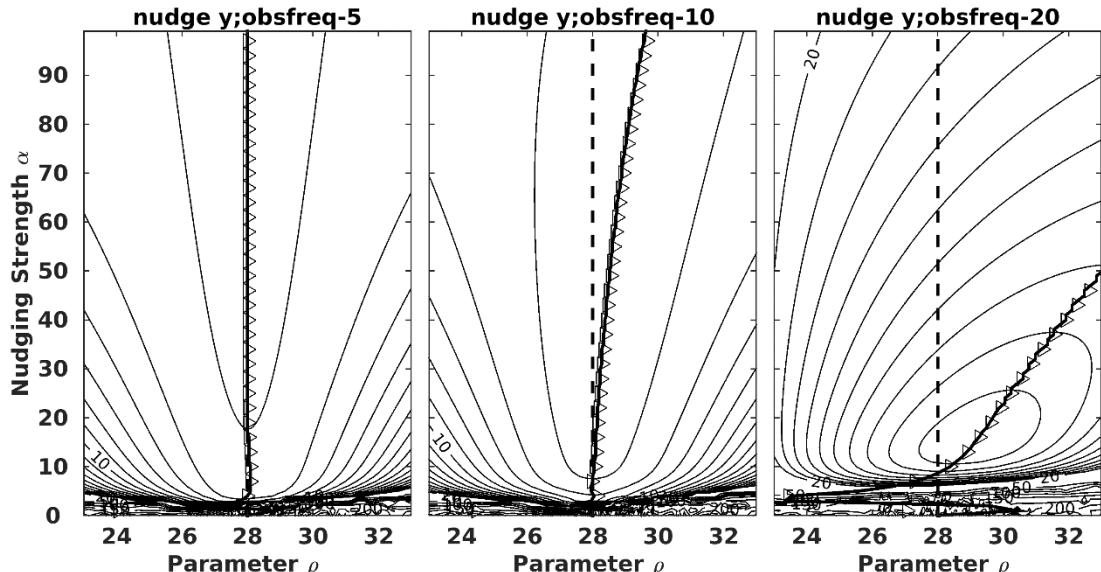


Figure 4.7. Cost functions depending on the parameter  $\rho$  and the nudging strength  $\alpha$  with nudging y for an observation frequency of every 5 steps, 10 steps and 20 steps as indicated by the titles. The thick black lines represent the global minimum with different nudging coefficients, the black triangles are the estimated parameter  $\rho$  after the optimization.

In real implementations, observations which usually come from reanalysis datasets are provided at a fixed time interval. Linear interpolation is used to fill the gaps between two-time instances which may force the system to an unphysical linear evolution. In this part, the influence of the observation frequency on the parameter estimation will be investigated. Observations are provided every 5 steps, 10 steps and 20 steps. The sets of variables used for nudging is y. Results of the other sets of variables used for nudging are similar to nudging y except for nudging xz and nudging x which do not lead to synchronization with the observations for 10-step and 20-step gaps. Expt4n is repeated by gradually increasing the nudging strength from 0 to 100 with observations every 5 steps, 10 steps, 20 steps. Linear

interpolation is used when nudging between two observational time steps. The assimilation window is still 20 TUs.

As Figure 4.7 shows, although all secondary minima are avoided with increasing nudging strength, the global minima of the slave system shown as the thick black lines are shifted, and the optimization reaches the global minima by tuning the parameter as the black triangles show. This shift is more serious with sparser observations than denser observations. With interpolation, linear evolution is assumed between two observational time levels and forced on the slave system. The optimal nudging strength will introduce the least error and keeps the most nonlinearity of the system. The optimal nudging strength is also diagnosed with:  $\alpha = 18$  for 5 steps,  $\alpha = 17$  for 5 steps and  $\alpha = 14$  for 20 steps.

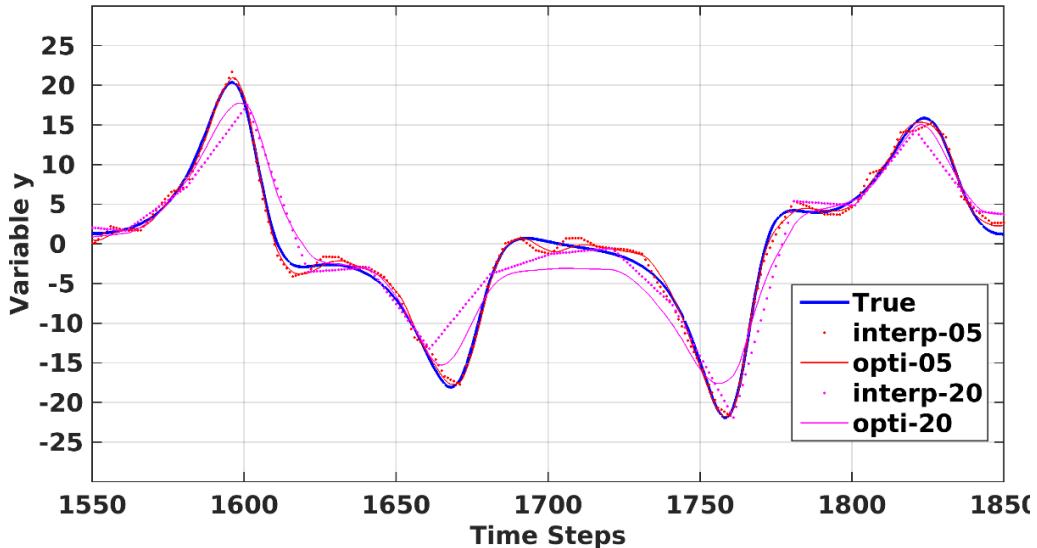


Figure 4.8. Trajectories of the variable  $y$ . The blue line is the true trajectory; the dotted lines are the linear interpolated trajectories with observations every 5steps (red dots) and 20 steps (magenta dots). The solid lines are trajectories after optimization with the optimal nudging strength for observations every 5steps (red line) and 20 steps (magenta line).

In Figure 4.8, we plot the interpolated  $y$  variable used for nudging and the  $y$  trajectory after optimization with the optimal nudging strength for observations every 5 steps and every 20 steps. For the interpolation between 5 steps which is much smaller than the predictability time of the Lorenz system, the slave system still follows the true trajectory reasonably well. Although some of the extreme values (at around 1825 steps) and transitions( $y \approx 0$ ) are not very well captured in the interpolated observations, the nudging term helps the slave system to better follow the nonlinear evolution without introducing unwanted transitions. With observations every 20 steps, the interpolated trajectory is stepwise. With its optimal nudging strength, the nudging term still recovers some of the nonlinear evolution of the system although not as good as observed every 5 steps. But when applied the strong nudging term, the slave system will follow the interpolated trajectory without catching the nonlinear transition points. Linear evolution will dominate the slave system which is different from the dynamics of observation system. The optimization will tune the parameter  $\rho$  to reduce the cost

function in this improper stepwise linear system which causes the estimated parameter  $\rho$  to significantly shift away from the true value.

Three groups of experiments are performed with the optimal nudging strength and nudging  $y$  for observations every 5 steps, 10 steps and 20 steps. The assimilation window is gradually extended from 2TUs to 300 TUs to reduce the compensation effect of a low number of noisy observations.

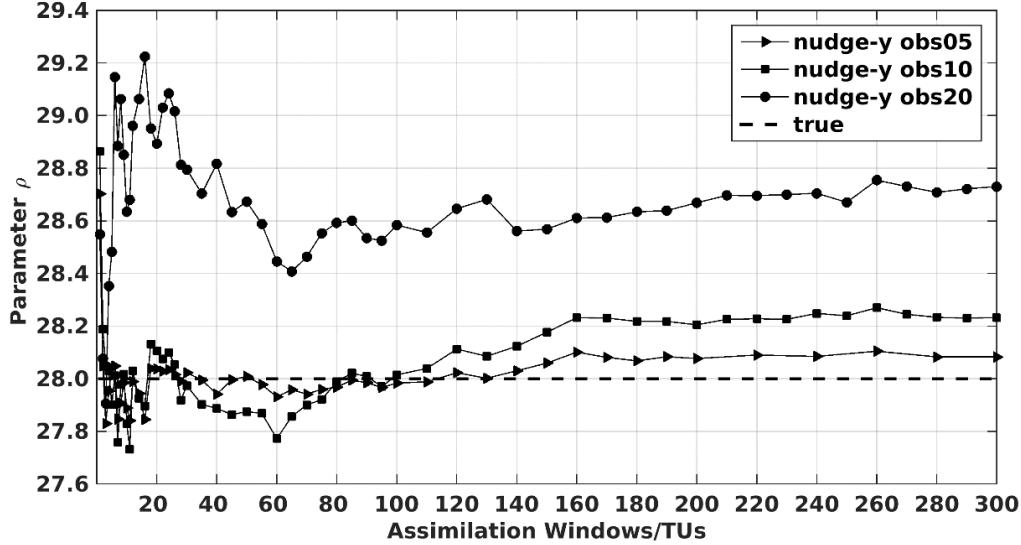


Figure 4.9. The estimated parameter  $\rho$  depending on the assimilation window for nudging  $y$  with the optimal nudging strength for different observation frequencies (see legend).

Figure 4.9 shows the estimated parameter  $\rho$  depending on the size of the assimilation window for different observation frequencies. Similar to Figure 4.6, the parameter estimation biases due to a low number of noisy observations are visible when the assimilation window is small. The case with observations every 20 steps shows stronger compensation than the cases with observations every 5 steps and 10 steps within 40 TUs, because the number of error realizations is low and the mean effect is correspondingly larger. By extending the assimilation window, the parameter estimation of  $\rho$  is robust among all 3 cases. The case with observations every 5 steps achieves the best estimation while the case with observations every 20 steps achieves the worst. The difference is caused by different synchronization efficiency with different observation frequencies. Overall, by using the optimal nudging strength, we can still reasonably and robustly estimate the parameter  $\rho$  with the given sparse observations.

In summary, parameter estimation based on this simplified DSPE method depends on the synchronization efficiency. Observational noise degrades parameter estimation because it unavoidably will be introduced to the slave system which degrades synchronization. Variables chosen for nudging influences the parameter estimation because of their different synchronization efficiency. A low observation frequency has a more serious impact on this method because nudging will force a linear evolution of the slave system which makes the slave system significantly different from the nonlinear dynamics governed by the equations. By choosing an optimal nudging strength which best recovers the error-free observations, the

negative effect of noise and linear interpolation can be reduced, and the parameter  $\rho$  can be reasonably estimated.

### 4.3 Discussions and conclusions

Due to the chaotic nature of the nonlinear model, the feasible assimilation window of the variational assimilation method is limited by the predictability of the dynamical system. (Abarbanel et al., 2010; Abarbanel et al., 2009) introduced the DSPE method to determine unknown parameters and the unobserved model state from observed time series of the nonlinear system using long assimilation windows. Synchronization is used to regularize the nonlinear instability and to avoid the occurrence of second minima in the cost function. They advocated that the necessary nudging term can be penalized and brought down to zero during the optimization.

In this study, we further investigate the viability of the DESP method on parameter estimation with long assimilation windows. We find that the penalty term for nudging coefficient plays a critical role for the development of the nudging coefficient during the optimization, and hence significantly influences the success of the parameter estimation. A large penalty term is likely to push the nudging term below a threshold value, and the slave model becomes chaotic again, which stalls the parameter estimation. Under the presence of observation noise, a small or non-existing penalty term leads to the increase of the nudging strength because the noise can only be fitted via nudging. A considerable large nudging coefficient is obtained, and its size is controlled by the weight of penalty term of the nudging coefficient. Although the forward model is changed by adding nudging terms, the influence of the parameter errors remains due to its direct influence on the model simulation, which enables the retrieval of the correct parameters in the presence of nudging terms. To guarantee the success of optimization, the nudging strength should be strong enough to avoid transitions relative to the observed state caused by erroneous parameters or initial errors. In this case, the assimilation window can be extended to infinite time.

The presence of observation noise degrades synchronization to the true (noise-free) system, the estimated parameter  $\rho$  shifts from the true value for strong nudging. With increasing nudging coefficients, more noise will be introduced into the slave system. By using the optimal nudging strength which best recovers the error-free signal from the noisy observations, we can still accurately estimate parameter  $\rho$ . Variables chosen for nudging also influence the accuracy of parameter estimation due to their different synchronization efficiency. For real implementations, we suggest adding nudging terms for all prognostic equations, if the necessary dimensions are not known. As the parameter estimation benefits from variables chosen for nudging, other synchronization methods such as the delay-coordinate nudging (Pazó et al., 2015) may also improve the accuracy of the parameter estimation.

For sparse observations, linear interpolation is used to fill the gaps between observation times. Strong nudging strength will make the slave system follow the linear evolution which violates the nonlinear dynamics of the governing equations, and for large gaps, the violation leads to increasing biases in the estimated parameter. Optimal nudging strengths can also be

found which introduce the least noise while best following the nonlinear evolution. The optimal nudging strength doesn't vary too much for our selected observation frequency.

Our results show that the performance of the parameter estimation with this simplified DSPE method depends on the synchronization efficiency. The more effective the synchronization, the more accurate the estimated parameter will be. With given observations, optimal nudging strength should be predefined for a better parameter estimation. Practically, the nudging strength must be large enough to push the positive Lyapunov exponents to negative values.

This simplified DSPE method only needs small changes of the traditional adjoint method and is expected to be suitable to systematically tune the model's climatology by optimizing process parameters for coupled climate models. However, we have explored this simplified DSPE method only for the simple Lorenz (1963) model. For complex atmospheric models, there are a number of issues that we have not investigated here due to the simplicity of the Lorenz (1963) model. These issues are:

- “On-off” switches as discussed in (Mu and Wang, 2003; Xu, 1996; Zou, 1997). Zero-order discontinues processes such as convective rain will directly degrade the usefulness of the tangent linear approximation because in the optimization we use gradients to make predictions for the effect of finite perturbations for which the gradients are no longer a good approximation. Some modification such replacing the “on-off” by smooth functions (Knorr et al., 2010; Županski and Mesinger, 1995) can be used to reduce the discontinuities in the forward model.
- For real implementation with a coupled climate model, the ERA-Interim (Dee et al., 2011) and NCEP reanalysis (Kanamitsu et al., 2002), which provide all variables at a 6-hour interval, could be used for synchronization. It is not clear whether the frequency is sufficient, but technique like delay-coordinate nudging (Pazó et al., 2015) may be employed to enhance synchronization. Moreover, further complications arise from additional model errors not described by the optimized parameters that project on the parameter and impact the parameter estimation.

Overall, it is reasonable to extend the feasible assimilation window for parameter estimation by using nudging terms to regularize the fast-growing unstable modes. In next section, the method will be further applied to a more complex earth simulation system, the Planet Simulator.

# Chapter 5

## The Climate model and configurations

Although the Lorenz (1963) system used in chapter 4 provides an interesting analog for Rayleigh-Bénard convection, it lacks many characteristics of a full, three-dimensional atmospheric system. For further tests on the usefulness of this simplified DSPE method on the parameter estimation, the Planet Simulator (PlaSim, Fraedrich et al., 2005a) is used here. Its exact adjoint model has been generated by an automatic differentiation tool TAF (Giering and Kaminski, 1998) and initially tested by Blessing et al. (2014) which provide us a good test base for the modified DSPE method. In this chapter, we briefly introduce this model and the parameterization processes that will be used in parameter estimation.

### 5.1 The model and its configurations

The dynamic core of PlaSim is the Portable University Model of the Atmosphere(PUMA) (Fraedrich et al., 2005a) including parameterization schemes for radiation, diagnostic cloud cover, large scale precipitation, convective precipitation, and dry convective adjustment. PUMA uses a spectral transform method horizontally, finite differences in the vertical and a semi-implicit time step. Other modules are reduced to linear systems including a thermodynamic sea-ice module, a land surface component and terrestrial biosphere component (SIMBA), a slab ocean module. Detailed description of this model and its parameterization schemes can be found online<sup>2</sup>.

The model resolution used in this study is  $5.6^\circ$  (T21) with ten vertical sigma levels for the atmosphere. The time step is 45 minutes. In our study, two configurations are used: (1) the “maximal” configuration which uses most of PlaSim’s component except for the SIMBA module and the slab ocean module is replaced with climatological surface temperature; (2) the “minimal” configuration which excludes the hydrological cycle and use climatology land surface and sea surface temperature to replace the correspondent components. For both configurations, the model is spun up for ten years using monthly averaged climatological sea ice concentration sea surface temperature (SST), glacier coverage, albedo and soil temperature. The restart files at the end of the spin-up are stored and used for all identical twin experiments below. For assimilating ERA-Interim data, the model is forced by 6-hourly ERA-Interim surface temperature rather than the climatology.

The “minimal” configuration is much more linear and simpler than the “maximal” configuration due to excluding the parameterization of moisture processes. For the minimal setup, Blessing et al. (2014) showed that the parameter could be accurately estimated after the optimization. The “minimal” configuration is used here to diagnose potential problems when linearizing the vertical diffusion parameterization schemes. We will also test the performance of the simplified DSPE with long assimilation window based on identical twin experiments.

---

<sup>2</sup><https://www.mi.uni-hamburg.de/en/arbeitsgruppen/theoretische-meteorologie/modelle/plasim.html>.

For assimilating the ERA-Interim reanalysis data (Dee et al., 2011), the “maximal” configuration is needed and therefore testing is done before that. The usefulness of the TAF generated tangent linear model and adjoint model will be tested step-by-step. And identical twin experiments will also be performed to investigate the performance of the modified DESP method.

Figure 5.1 (a) and (b) show the annual mean of zonal averaged zonal wind and temperature for the ERA-Interim data and the PlaSim simulation (“maximal” configuration). The model simulation matches observed atmosphere structures well. For the zonal mean of the eastward component of the wind  $u$ , westerlies (flow from west to east in meteorology) prevail almost everywhere except close to the equator where easterlies prevail. The westerlies peak in the subtropical jet stream which is centered at around 30-40 degree latitude and stronger in the southern hemisphere than in the northern hemisphere. But the peak speeds of westerlies in the PlaSim simulation are stronger than ERA-Interim for both hemispheres.

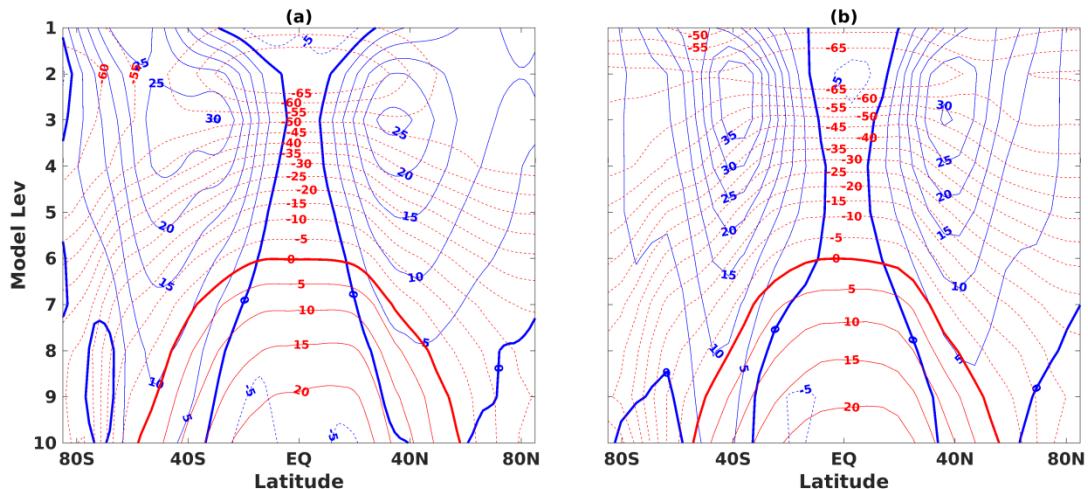


Figure 5.1. Annual mean of zonal averaged zonal wind (m/s) and temperature ( $^{\circ}$ C) for the ERA-Interim reanalysis (a) and the PlaSim simulation (b). Annual mean of the ERA-Interim data use data of year 1992.

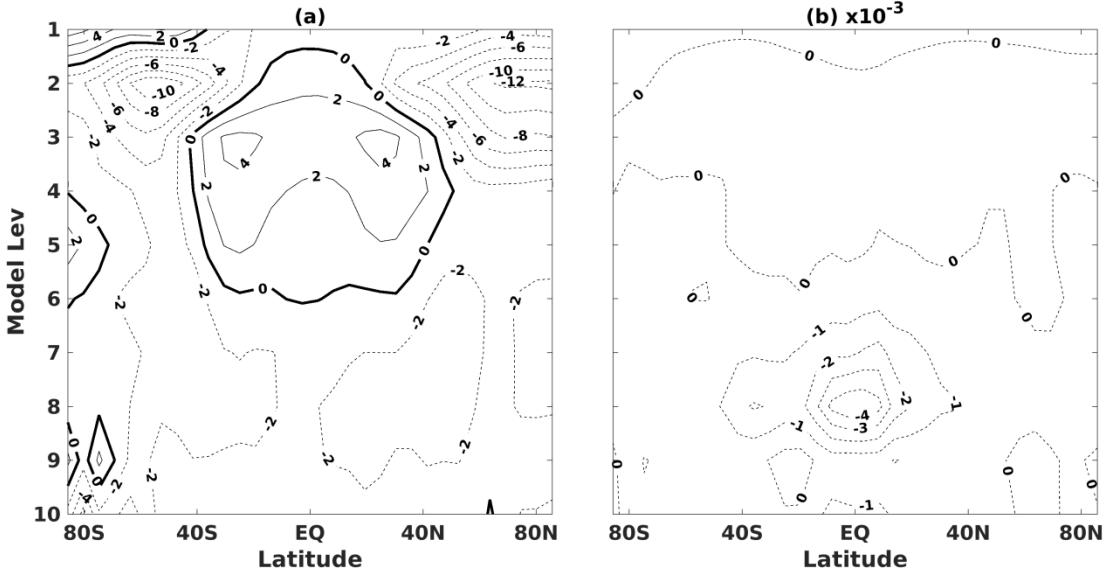


Figure 5.2. Zonal averaged climatological temperature difference ( $^{\circ}\text{C}$ ) (a) and specific humidity difference ( $\text{kg}/\text{kg}$ ) (b) between the PlaSim simulation and the ERA-Interim reanalysis. The difference is computed by the PlaSim simulation minus ERA-Interim reanalysis. Annual mean of ERA-Interim data uses data of year 1992.

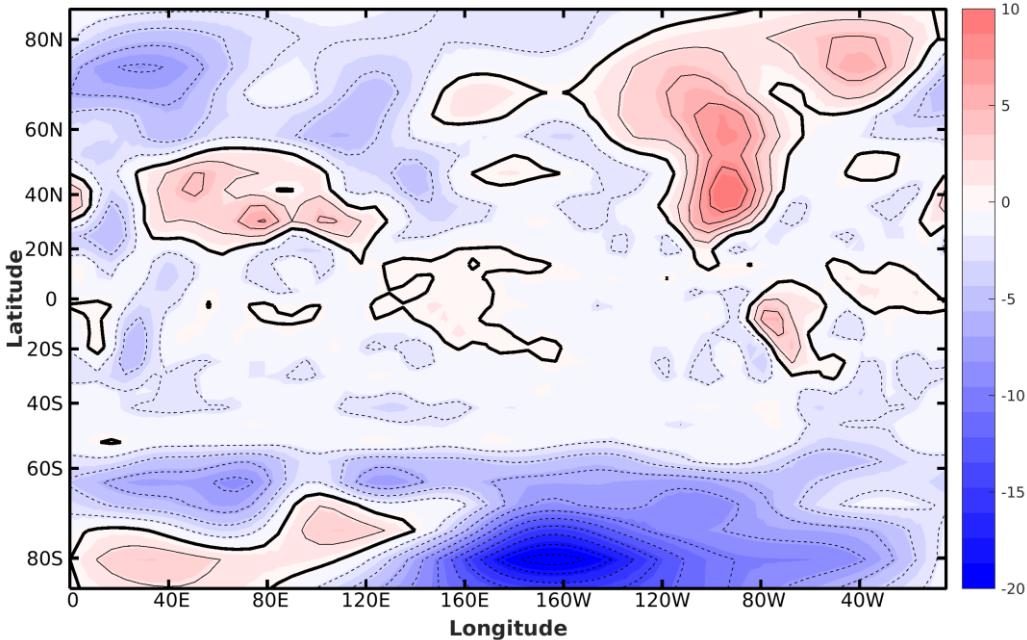


Figure 5.3. Temperature difference in level 10 between the PlaSim simulation and the ERA-Interim reanalysis for model level 10. The difference is computed by the PlaSim simulation minus ERA-Interim data. Annual mean of the ERA-Interim data uses data of year 1992. The contour interval is  $2\text{ }^{\circ}\text{C}$ .

For air temperature, the PlaSim simulation shows a cold bias almost everywhere except for the equator and the subtropical region at the top of the atmosphere as in Figure 5.2 (a). An

average of -2 °C bias is visible for the lower atmosphere and much larger (around -4 °C) in the Antarctic. At the top of the atmosphere, a warm bias with amplitudes around 2-4 °C exists in the tropics and subtropical region while a cold bias with amplitudes larger than -10 °C dominates the polar region. At the topmost level of the Antarctic region, a warm bias exists. Figure 5.3 shows the temperature difference for level 10. A warm bias exists over Asia, northern America, and Antarctica from 0E to 120E. A cold bias dominates all other regions. For specific humidity, the PlaSim simulation has less water vapor (~20% less than the ERA-Interim reanalysis) in the whole atmosphere as Figure 5.2 (b) shows.

## 5.2 Description of physical parameterizations

For the parameter estimation, a couple of physical parameterizations which include uncertain process parameters are selected. The first ten parameters have been tested in the previous study (Blessing et al., 2014) based on identical twin experiments. In this study, six more parameters are included. We briefly introduce these physical processes and the parameters. The parameters, their default values and possible ranges are listed in table 5.1. The first nine parameters are related to the dynamical core of the model and used in identical twin experiments with the “minimal” configuration. The rest of the parameters are related to the radiation parameterization, and tune cloud characteristics and water vapor absorptivity of longwave radiation. Four of them together with the first nine parameters are used in identical twin experiments with the “maximal” configuration while all of them are tested in assimilating the ERA-Interim data.

Table 5.1. Process parameters used in parameter estimation

Parameter	Description	Default	Range
<i>tfrc1</i>	time scale for Rayleigh friction (lev01)	20 days	>0 days
<i>tfrc2</i>	time scale for Rayleigh friction (lev02)	100 days	>0 days
<i>tdissd</i>	diffusion time scale for divergence	0.2 days	>0 days
<i>tdissz</i>	diffusion time scale for vorticity	1.1 days	>0 days
<i>tdisst</i>	diffusion time scale for temperature	5.6 days	>0 days
<i>vdiff_lamm</i>	Parameter in vertical diffusion	160 m	0-1000m
<i>vdiff_b</i>	Parameter in vertical diffusion	5 m	>0 m
<i>vdiff_c</i>	Parameter in vertical diffusion	5 m	>0 m
<i>vdiff_d</i>	Parameter in vertical diffusion	5 m	>0 m
<i>tpofmt</i>	Tuning of point of mean transmittivity in layer	1.	0-1
<i>tswr1</i>	Tuning of cloud albedo range1( $\lambda < 0.75 \text{ um}$ )	0.04	0.01-0.2
<i>tswr2</i>	tuning of cloud back scattering range2( $\lambda > 0.75 \text{ um}$ )	0.048	0.01-0.5
<i>tswr3</i>	tuning of cloud scattering albedo range2( $\lambda > 0.75 \text{ um}$ )	0.004	0.001-0.1
<i>th2oc</i>	absorption coefficient water vapor continuum	0.04	0.005-0.1
<i>acllwr</i>	mass absorption coefficient for clouds	0.1	>0
<i>tdissq</i>	diffusion time scale for specific humidity	0.2 days	>0days

### 5.2.1 Rayleigh friction

A Rayleigh friction scheme is used in the uppermost two levels to represent gravity wave drag effects at the top of the atmosphere. Rayleigh friction is done in spectral space, and its contribution to the tendency of divergence or vorticity X is:

$$\frac{\partial X}{\partial t} = -\frac{1}{\tau} X \quad (5.1)$$

where  $\tau$  is the timescale (days) for the Rayleigh friction. In this study,  $\tau$  for the uppermost two levels are used in parameter estimation.

### 5.2.2 Horizontal diffusion

The horizontal diffusion parameterization bases on Laursen and Eliasen (1989). The diffusion is also done in spectral space. Its contribution to the tendency of temperature, divergence, vorticity and moisture X is:

$$\frac{\partial X_n}{\partial t} = -k_x L_n X_n \quad (5.2)$$

where n is the wave number and  $L_n$  is a scale selective function of the total wave number and acts to enhance damping effects with increasing wave number n:

$$L_n = \begin{cases} (n - n_*)^\alpha & \text{for } n > n_* \\ 0 & \text{for } n < n_* \end{cases} \quad (5.3)$$

The cut-off wave number is 15 and  $\alpha=2$ . The diffusion time scales  $k_x$  for temperature ( $tdissst$ ), divergence ( $tdissd$ ), vorticity ( $tdissz$ ) and specific humidity ( $tdissq$ ) are used in parameter estimation and their default value is listed in table 5.1.

### 5.2.3 Vertical diffusion parameterization

The unresolved turbulent exchange is represented by vertical diffusion parameterization scheme based on Louis (1979). The calculation of the exchange coefficient  $K_m$  and  $K_h$  are as follows:

$$K_m = l_m^2 \left| \frac{\partial \vec{v}}{\partial z} \right| f_m(R_i) \quad (5.4)$$

$$K_h = l_h^2 \left| \frac{\partial \vec{v}}{\partial z} \right| f_h(R_i) \quad (5.5)$$

where  $l_m$  and  $l_h$  are the mixing lengths (Blackadar, 1962) and defined as:

$$\frac{1}{l_m} = \frac{1}{kz} + \frac{1}{\lambda_m} \quad (5.6)$$

$$\frac{1}{l_h} = \frac{1}{kz} + \frac{1}{\lambda_h} \quad (5.7)$$

where  $\lambda_h = \lambda_m \sqrt{3d/2}$ . The default values are:  $\lambda_m=160$  and  $d=5$ . The functions  $f_h(R_i)$  and  $f_m(R_i)$  are empirical function of Richardson number based on Monin-Obukhov similarity theory.

For stable conditions ( $R_i \geq 0$ ),  $f_m(R_i)$  and  $f_h(R_i)$  are given by:

$$f_m = \frac{1}{1 + (2bR_i/\sqrt{1+dR_i})} \quad (5.8)$$

$$f_h = \frac{1}{1 + (3bR_i/\sqrt{1+dR_i})} \quad (5.9)$$

while for the unstable cases ( $R_i < 0$ ):

$$f_m = 1 - \frac{2bR_i}{1 + (3bc\frac{l_m^2}{\Delta z^{3/2}z^{1/2}}[(\frac{z+\Delta z}{z})^{1/3}-1]^{3/2}/\sqrt{-R_i})} \quad (5.10)$$

$$f_h = 1 - \frac{3bR_i}{1 + (3bc\frac{l_m^2}{\Delta z^{3/2}z^{1/2}}[(\frac{z+\Delta z}{z})^{1/3}-1]^{3/2}/\sqrt{-R_i})} \quad (5.11)$$

where b, c are constant values and their default values are: b=5 and c =5. For this parameterization scheme, the uncertain parameters to be tuned are  $\lambda_m$  (*vdiff\_lamm*), b (*vdiff\_b*), c (*vdiff\_c*) and d (*vdiff\_d*).

#### 5.2.4 Radiation parameterization

A diagnostic cloud parameterization scheme based on Slingo and Slingo (1991) is used in this model. Five parameters related to the cloud properties, which influence shortwave radiation and longwave radiation, are chosen for parameter estimation. Three parameters are related to the transmissivity of clouds for solar radiation.  $f_{b1}$  is related to backscatter of solar radiation with wavelength  $\lambda < 0.75$  um and takes the form:

$$T_{C1} = \frac{1}{1 + f_{b1}\tau_{N1}/\sqrt{u_0}} \quad (5.12)$$

where  $\tau_{N1}$  is effective optical depth based on Stephens (1978).  $u_0$  is the solar zenith angle.  $f_{b1}$  (*tswr1*) is a tunable parameter, which comprises a tuning opportunity for cloud albedo, and is set to a default value of 0.03335.

Parameters  $f_{b2}$  (*tswr2*) and  $f_{o2}$  (*tswr3*) are used to approximate the table values of the backscatter coefficients  $\beta_2$  and  $\omega_0$ :

$$\beta_2 = \frac{f_{b2}\sqrt{u_0}}{\ln(3+0.1\tau_{N2})} \quad (5.13)$$

$$\omega_0 = 1 - f_{o2}u_0^2\ln(1000/\tau_{N2}) \quad (5.14)$$

where  $\tau_{N2}$  is the optical depth based on Stephens (1978). The default values of  $f_{b2}$  (*tswr2*) and  $f_{o2}$  (*tswr3*) are 0.048 and 0.004.

Parameters  $k_{cl}$  (*acallow*) and  $k_{h2o}$  (*th2oc*) are used to parameterize absorption effects of clouds and water vapor on longwave radiation and the absorptivity are given by:

$$A^{cl} = 1 - \exp(-\beta_d k_{cl} W_L) \quad (5.15)$$

$$A^{h2o} = 1 - \exp(-k_{h2o} W_{L1}) \quad (5.16)$$

where  $W_L$  is the cloud liquid water path and  $W_{L1}$  is liquid water path. The default values are:  $k_{cl} = 0.1$  and  $k_{h2o} = 0.25$ .

The last parameter chosen for parameter estimation is  $f_T(t_{pofmt})$  which provides a tuning opportunity for transmissivity of a layer partially filled with clouds and is given by:

$$T(l, l - 1) = f_T T_{(l, l - 1)}^{cs} (1 - cc_{(l, l - 1)} A_{(l, l - 1)}^{cl}) \quad (5.17)$$

where  $T_{(l, l - 1)}^{cs}$  is the clear sky transmissivity and  $A_{(l, l - 1)}^{cs}$  is the cloud emissivity. Its default value is set to  $f_T=1$ .

### 5.3 Implementation of nudging terms

For the implementation of the modified DSPE method, simple nudging terms (Hoke and Anthes, 1976) are added to the climate model. For PlaSim, a spectral transform method is used for solving the primitive equations. Model integration including 4 steps: 1) compute the tendencies caused by adiabatic nonlinear term effects in physical space; 2) compute the diabatic tendency in spectral space and adding the adiabatic perturbation; 3) compute the parameterization effects including: radiation, precipitation, planet boundary etc. in physical space; and 4) compute damping effects in spectral space. The nudging terms are added to the tendencies in spectral space after step 4) with nudging divergence, vorticity, temperature. Moisture is also included in the nudging when the “maximal” configuration is used. The implementation of the nudging terms as the spectral coefficient of the model state is as follows:

$$\frac{\partial X_n}{\partial t} = -\frac{1}{\tau_{relax}} (X_n - X_n^{obs}) \quad (5.18)$$

where  $\tau_{relax}$  is a nudging timescale. In this study, an implicit time step scheme is used and the equation (5.18) is rewrite as:

$$\frac{X_n^{t+1} - X_n^t}{\Delta t} = -\frac{1}{\tau_{relax} + \Delta t} (X_n^t - X_n^{obs}) \quad (5.19)$$

### 5.4 Summary

In this chapter, we provided details of PlaSim and the configurations that will be used for the assimilating experiments. For the “maximal” configuration, we see that the simulation with PlaSim resembles basic characteristic of the ERA-Interim data. But PlaSim simulates much stronger westerlies than the ERA-Interim data, and the air temperature is colder except for the upper layer in the tropics. PlaSim also has less moisture than the ERA-Interim data. It is difficult to identify the causes of these biases. The parameterization schemes used in this study are introduced. Some of the parameterization schemes have very complex forms, which makes it harder to code the TLM and adjoint model. But with automatic difference tools TAF, the TLM and adjoint model are automatically generated. In the study below, we will try to reduce model bias by tuning the introduced process parameters.

# Chapter 6

## Assimilation experiments with PlaSim

In this chapter, the modified DSPE method is tested with the “minimal” configuration and the “maximal” configuration of PlaSim. The usefulness of the TLM and the adjoint model is tested by comparing with the finite differences of the perturbed model integrations. Modifications are made to the TLM and the adjoint model to improve the description of the error propagation of the nonlinear model. We start with the “minimal” configuration which is more linear than the “maximal” configuration. Throughout this study, the original nonlinear system is called the free system or PlaSim while the system with nudging terms is called the slave system.

### 6.1 The cost function

The cost function used for identical twin experiments is defined as following:

$$J(p) = \sum_{k=1}^{lev} \frac{(T-T_{era})^2}{\sigma t_k^2} + \sum_{z=1}^{lev} \frac{(Vo-Vo_{era})^2}{\sigma t_k^2} + \sum_{k=1}^{lev} \frac{(D-D_{era})^2}{\sigma d_k^2} + \frac{(SP-SP_{era})^2}{\sigma_{sp}^2} \quad (6.1)$$

where T, Vo, D, and SP represent air temperature, vorticity, divergence and surface pressure. Specific humidity is not included in the cost function because: (1) there is no water vapor in the “minimal” configuration, and (2) for the “maximal” configuration, the tangent linear model cannot provide a reasonable approximation of the nonlinear perturbation of moisture due to the strong nonlinear moisture parameterizations, such as convective precipitation and dry convective adjustment. The observations are time averaged over the whole assimilation window to represent the mean state. For historical reason, the observations used in the “minimal” configuration don’t include air temperature for model level 1 (upper most level of the model).

### 6.2 Modifying the TLM and its adjoint

For PlaSim, the TLM and the adjoint model have been automatically generated by TAF. The correctness of the TLM and the adjoint model is validated in Blessing et al. (2008). For the “minimal” configuration, Blessing et al. (2014) also provided four examples in which the parameters are accurately retrieved with an assimilation window of 56 days and a reduced time step of 10 minutes. But with larger time steps (20min and 45min in their study), the optimization fails for such a long assimilation window. Although the simulations with time steps of 45 minutes and 10 minutes show some differences, it is unlikely that the differences change the predictability of the system and hence change the feasible assimilation window. It is likely that the accuracy of the TLM and its adjoint depends on the time step, due to strong nonlinearity parameterizations such as vertical diffusion parameterization. Although reducing the time step may improve the validity the TLM and the adjoint model (Blessing et al., 2014; Zhu and Kamachi, 2000), the computational cost is also greatly increased. Here, potential problems with the exact TLM and the adjoint model generated by TAF are further tested. Small modifications based on Zhu et al. (2002) are made on the exact TLM and the adjoint

model to improve the usefulness of the adjoint model without reducing the time step. The problem will be shown based on the “minimal” configuration and modifications are made when linearizing the vertical diffusion parameterization. Then assimilating experiments will be performed to validate the usefulness of the modified adjoint model. For this “minimal” configuration, the first nine parameters listed in Table 5.1 are used for the parameter estimation.

To validate the exact TLM generated by TAF, the time evolution of root mean square errors (RMSEs) for the exact TLM and a perturbed PlaSim integration with 0.0001% perturbation on all nine parameters are computed based on equation (6.2) and (6.3):

$$RMSEs = \sqrt{\frac{1}{N} (X_p - X_{ref})^2} \quad (6.2)$$

$$RMSEs = \sqrt{\frac{1}{N} (M(\Delta p))^2} \quad (6.3)$$

where  $X$  represents the model state used in the cost function, subscripts  $p$  and  $ref$  represent a perturbed integration and a reference integration.  $M$  is the TLM operator, and  $\Delta p$  is the parameter perturbation.  $N$  is the number of the model grid on one level (32x64).

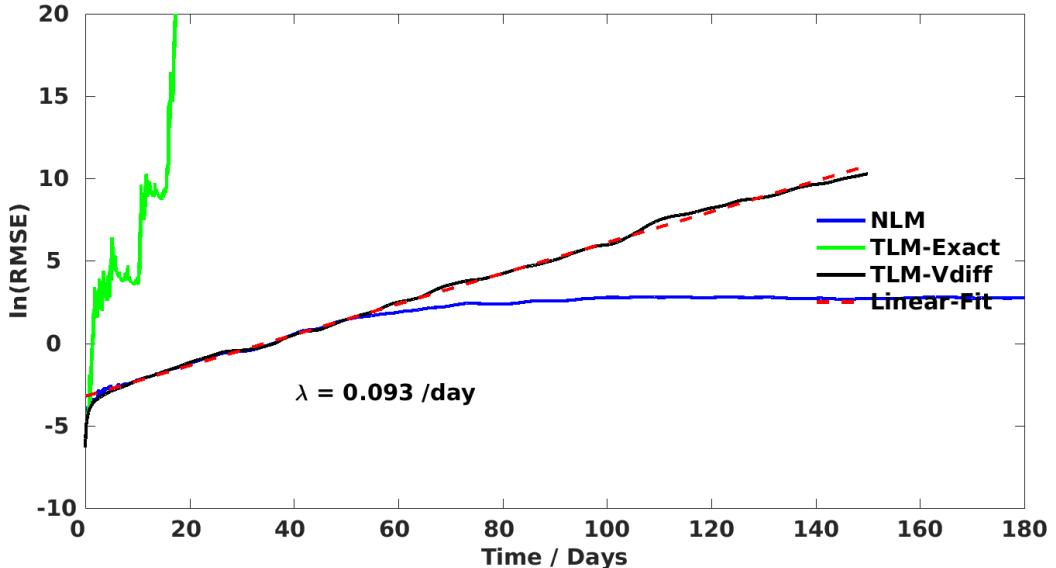


Figure 6.1. The time evolution of instantaneous RMSEs for the exact TLM (the green line), the modified TLM (the black line) and a perturbed PlaSim integration (the blue line) with a perturbation size of 0.0001%. The red dashed line is a linear fit of the blue line for the first 50 days.

The green line and blue line in Figure 6.1 show the time evolution of RMSEs from the exact TLM and a perturbed PlaSim integration. The leading local Lyapunov exponent  $\lambda=0.093$  /day is crudely computed based on a linear fit of the nonlinear error propagation within the first 50 days. The predictability limit of this “minimal” configuration is around  $\frac{1}{\lambda} \approx 10$  days and a feasible assimilation window should be around 20 days. The error propagation in the exact TLM matches that of the PlaSim quite well within one day. Beyond

one day, the RMSEs from the exact TLM grow dramatically and are totally different with that from PlaSim. For integration time larger than one day, the exact TLM cannot reasonably describe error propagation and hence the adjoint model will not provide useful gradient information that points to the global minimum. An explanation is that: the nonlinear model contains fast-growing modes, which saturate quickly and have little effects on the model simulation. Directly linearizing these processes will lead to a dramatical error growth rate in the TLM as the green line in Figure 6.1 shows and limit the usefulness of the TLM model and its adjoint. Regularizations are needed for improving the usefulness of the TLM and the adjoint model.

For the “minimal” configuration, the dramatically increase of RMSEs in the exact TLM is likely caused by linearizing the vertical diffusion parameterization. Previous studies (Errico et al., 1993; Janisková et al., 1999; Laroche et al., 2002; Mahfouf, 1999; Zhu et al., 2002) have shown that regularizations are needed when linearizing the vertical diffusion parameterization. One reason is that directly linearizing the vertical diffusion parameterization may lead to unstable of the TLM (conflicting with the CFL condition, Zhu and Kamachi (2000)). Another reason is that this parameterization contains fast-growing modes, but these modes saturate quickly and therefore have little effects on the model simulation. However, linearizing these fast-growing modes will lead to the dramatical error growth in the TLM. Among the several methods for regularization are: (1) neglect  $K'_m$  (Errico et al., 1993; Mahfouf, 1999); (2) smoothing the discontinue point in the forward model (Janisková et al., 1999; Laroche et al., 2002); (3) removing part of TLM which contribute to the instability of TLM while have little impact on the error propagation (Zhu et al., 2002). For the parameter estimation, the third method is adopted, which can retain most of the parameter information and which also has been tested to be useful by Liu et al. (2012).

In PlaSim, the vertical diffusion parameterization is based on Louis (1979) and is expressed as equations (5.4)-(5.11). For this parameterization scheme, the TLM will provide too large error propagation when the Richardson number changes signs (Janisková et al., 1999; Laroche et al., 2002). This indicates that the fast-growing modes in this parameterization are related to the Richardson number. The change of sign for the Richardson number represents the change of stability in the atmosphere. By try-and-error, we find that the TLM and the adjoint model can accurately represent the error propagation simply by setting  $Ri' = 0$ . The black line in Figure 6.1 shows the error propagation for the modified TLM which fits well to the nonlinear one until 50 days when nonlinearity acts to stop the linear error growth.

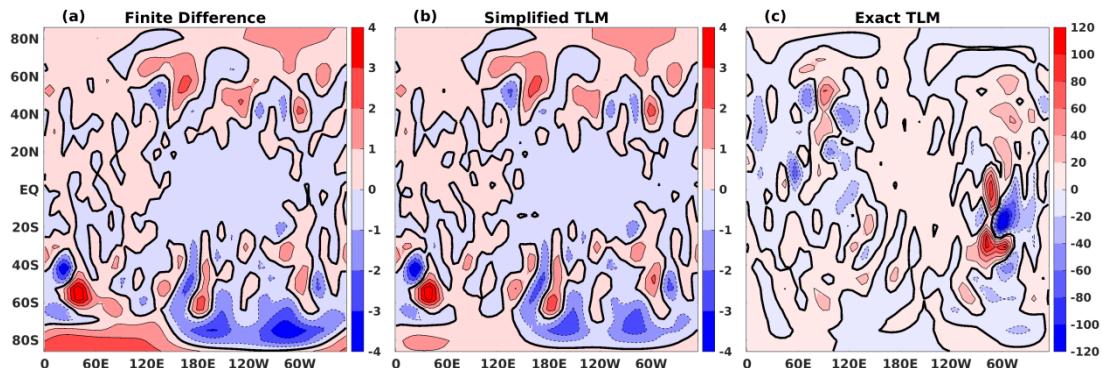


Figure 6.2. Spatial distribution of the surface pressure difference (10Pa) caused by 10% perturbations on all 9 parameters: (a) the finite difference between a perturbed PlaSim integration and the control integration, (b) the modified TLM, (c) the exact TLM. The integration time is 10 days.

Figure 6.2 plots the surface pressure difference between a perturbed PlaSim integration and the control integration (a), the modified TLM (b) and the exact TLM with 10% perturbations on all parameters for an integration of 10 days. Both the pattern and amplitude of the surface pressure difference are similar in the modified TLM and the finite difference which indicates for how long the linear dynamics dominate the nonlinear model and for how long the simplified TLM accurately represents error propagation of the nonlinear model. The exact TLM gives a much larger surface pressure difference than both the nonlinear model and the modified TLM. The results show that it is necessary to modify the TLM generated by TAF. By simply setting  $Ri' = 0$  when linearizing the vertical diffusion parameterization, the modified TLM can accurately describe the most relevant terms of the nonlinear error propagation.

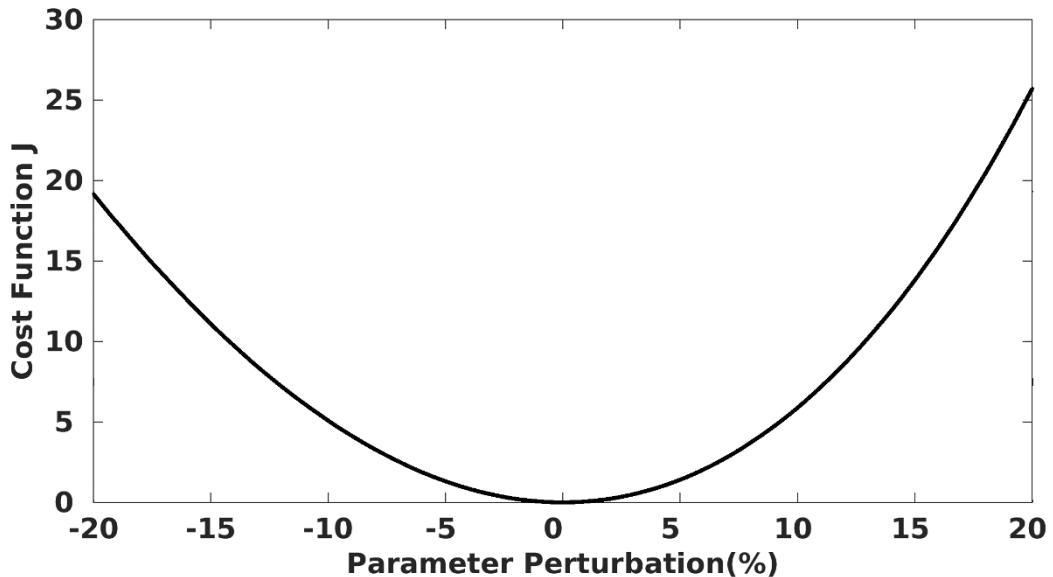


Figure 6.3. The cost function depending on parameter perturbations (%) for an assimilation window of 10 days. All 9 parameters are perturbed.

To test the usefulness of the modified adjoint model for parameter estimation, identical-twin experiments are performed with an assimilation window of 10 days. Figure 6.3 shows the cost function depending on the parameter perturbations. The cost function is quiet smooth and shows no chaotic behavior. From the cost function perspective, the optimization should converge if the adjoint model can provide useful gradients. Four assimilating experiments are performed which perturbed all nine parameters with perturbation sizes of 10%, -10%, 20%, -20%. The change of the cost function, the norm of gradients, and the norm of the parameter perturbations are plotted in Figure 6.4. For all four experiments, the norm of the parameter errors is reduced by more than three orders, and the cost function and its gradients are also strongly reduced. The true values of the parameters are successfully retrieved after the optimization.

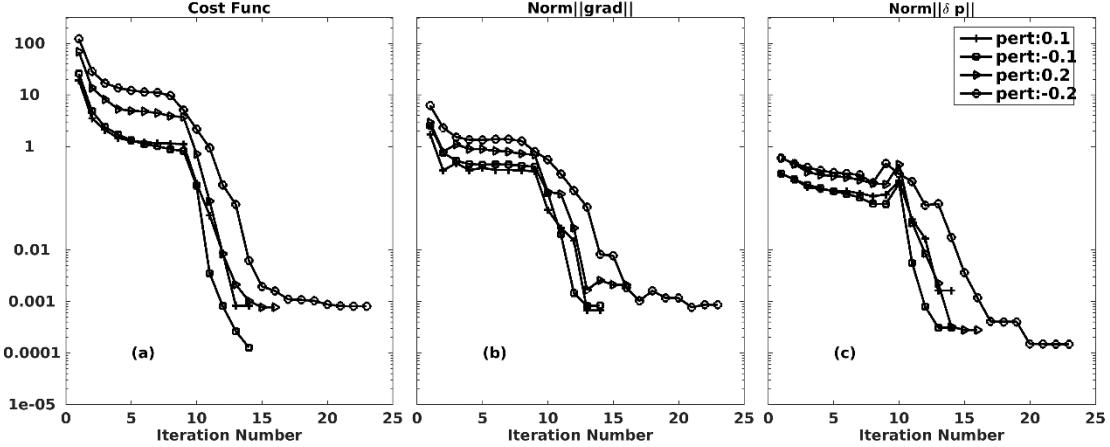


Figure 6.4. Convergence of the minimization for the minimal configuration for four experiments with different perturbation size (see legend) on all 9 parameters. The assimilation window is 10 days.

With the automatic differentiation tool TAF, the TLM and the adjoint model can be automatically generated, and the correctness can be verified through gradient check as in Blessing et al. (2008), which is usually done with very small perturbations and within a short integration period. While for data assimilation, the perturbation sizes are considerably larger and it is more practical to test the usefulness of the adjoint model as in Figure 6.2 shows. Although the TLM and the adjoint model are correct, their usefulness for data assimilation may be degraded due to linearize fast growing modes in the vertical diffusion parameterization. These modes contribute most to the unstable of the TLM but have little impacts on accurate representation of the TLM. As suggested by Zhu et al. (2002), this problem can be solved by setting  $Ri' = 0$ . Through performing identical twin experiments, we see that the modified adjoint model provides useful gradient information and the default values of the parameters are successfully retrieved within 30 iterations. In the following part, the simplified DSPE method will be tested with the modified adjoint model derived here. Nudging terms will be added to the forward model as in equations (5.18)-(5.19) and the assimilation window will be extended to 1 year.

### 6.3 Assimilating experiments with the “minimal” configuration

#### 6.3.1 The nudging strength for synchronization

To synchronize the chaos of the nonlinear system and observations which is required by the modified DSPE method, the nudging strength should be strong enough to push the positive Lyapunov exponent to negative values. For the “minimal” configuration, the nudging terms are added on air temperature (lev02-lev10), vorticity (all levels) and divergence (all levels) using the same nudging timescale. Based on the TLM, Figure 6.5 shows the logarithm of RMSEs with a 0.0001% perturbation on all 9 parameters for different nudging timescales. When decreasing the nudging timescales from 5 days to 2 days, the exponential increase of RMSEs gradually reduces and shows no visible exponential increase for a nudging timescale of 2 days. The nudging timescale is set to 2 days for this “minimal” configuration.

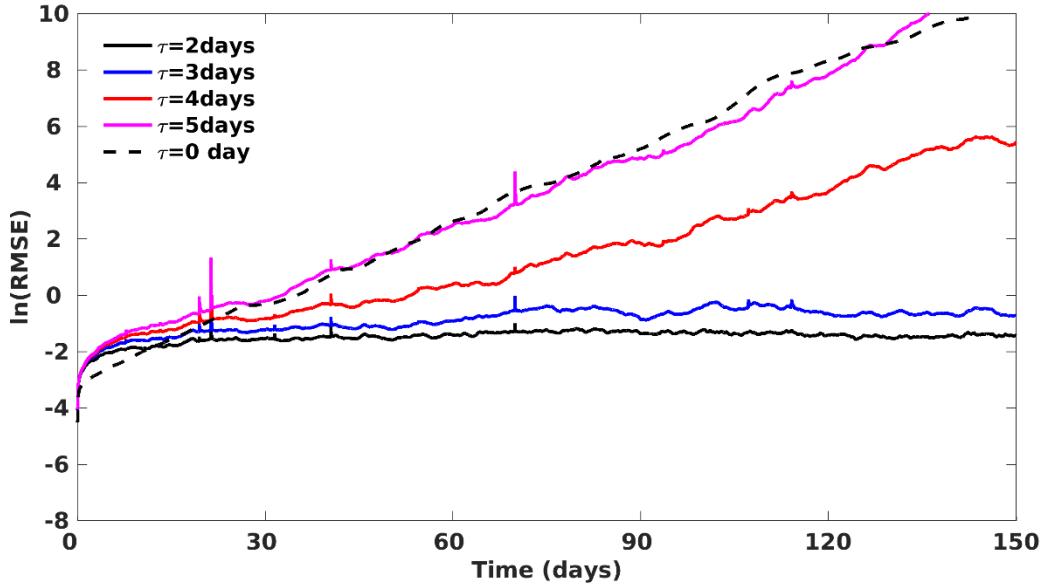


Figure 6.5. The time evolution of instantaneous RMSEs for the modified TLM without nudging ( $\tau=0$  day), with nudging timescales ranging from 5 days to 2 days (see legend).

Figure 6.6 shows the cost function components depending on the parameter perturbations ranging from -90% to 90% for the free system (the solid lines) and the slave system (the dashed lines) with a nudging timescale of 2 days for an integration time of 1 year. Without nudging terms, there are narrow valleys around the global minima which show strong sensitivity. Beyond this region, multiple minima occur, and the sensitivity is much smaller. Also, we see that the cost function is not symmetric which reflect the nonlinearity of the free system. With nudging terms, the cost function is smooth for all components. The cost function component for temperature and divergence is not strictly symmetric. Negative perturbation side shows larger gradient than the positive perturbation side. With nudging timescales smaller than 2 days, the slave system is non-chaotic, but nonlinearity still exists.

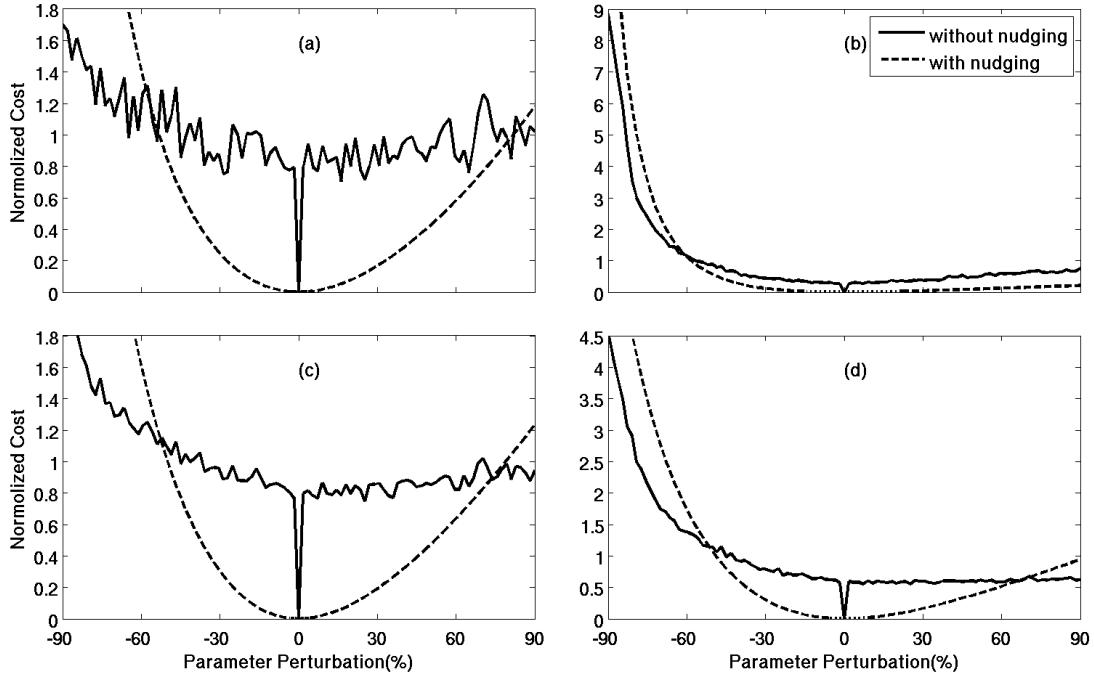


Figure 6.6. Cost function components depending on the size of parameter perturbations for PlaSim (the solid lines) and the slave system (the dashed lines): (a) surface pressure, (b) temperature, (c) vorticity and (d) divergence. The integration time is 1 year. The cost function is normalized by the mean value over all perturbations.

The TLM should provide error propagations that is similar to that from the finite difference between a perturbed and the control slave system integrations. Although the highly nonlinear processes in the moisture parameterization are excluded in this “minimal” configuration, there still exists some “if” and “where” statements (such as the vertical diffusion parameterization). These conditionals potentially degrade the usefulness of the TLM and its adjoint because that a small perturbation on the nonlinear model integration may change the trajectory shift to another side of the conditionals and the TLM based on the unperturbed trajectory cannot describe this. The problem is likely to be more serious with longer integration time because the conditionals are likely to take effects more frequently. For the slave system, the usefulness of TLM model is validated before performing assimilating experiments. This is done by comparing model state differences with a 10% perturbation on all nine parameters from the finite difference of perturbed slave system integrations and the TLM of the slave system after 1-year integrations.

Figure 6.7 shows the surface pressure difference of the slave system from the finite difference (a) and the TLM (b) after 1-year integration. The TLM fits well with the finite difference for both pattern and amplitude. Slightly differences are visible near Antarctica where the finite difference shows larger differences than the TLM. Overall, the conditionals in the “minimal” configuration don’t degrade the usefulness of TLM. With a 2-day nudging timescale, this slave system becomes non-chaotic, and the TLM can very well describe the error propagation of this slave system. Although the conditionals may slightly degrade the performance of the TLM, the TLM generated by TAF and modified by setting  $Ri' = 0$  can

provide an accurate description of the parameter sensitivity compared with the finite difference.

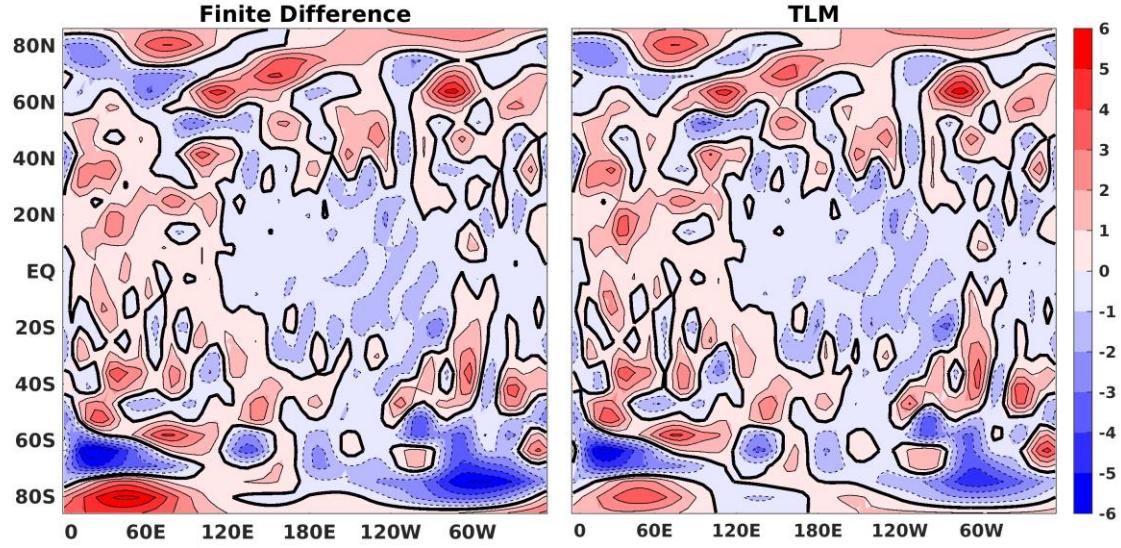


Figure 6.7. The surface pressure difference (Pa) caused by a 10% perturbation on all 9 parameters for the finite difference of perturbed slave system integrations (a) and the TLM of the slave system (b) after 1-year integrations.

### 6.3.2 Assimilation experiments

Three identical twin experiments are performed to investigate how the DSPE method works with the “minimal” configuration. The nine parameters are perturbed with sizes of 10%, 20% and -90% and the assimilation window is 1-year.

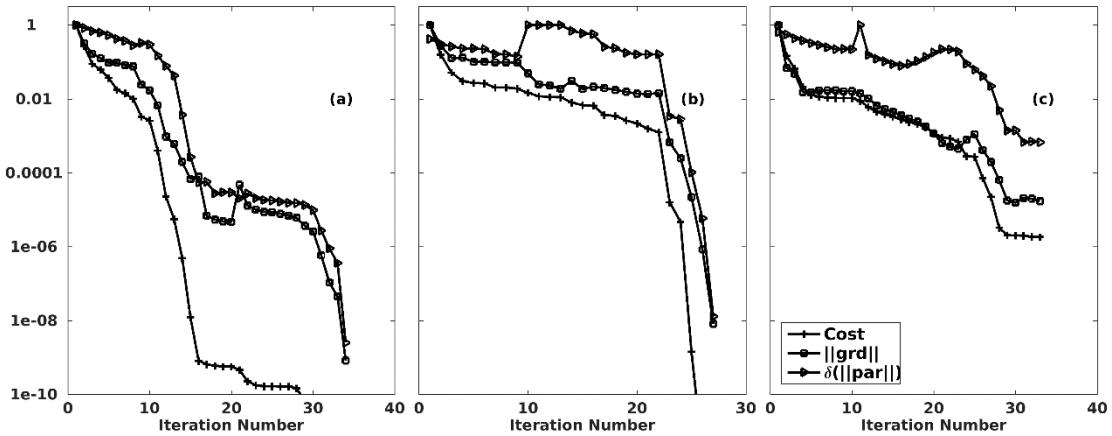


Figure 6.8. Convergence of the minimization for 3 assimilating experiments with an assimilation window of 1 year: (a) 10% perturbations (b) 20% perturbations (c) -90% perturbations. The cost function, norm of gradients and norm of parameter perturbation shown in the figure are normalized by their maximum values, respectively.

Figure 6.8 gives convergence of the minimization for the three experiments. For all three experiments, the norm of the parameter perturbations is reduced by more than three orders

with a significant reduction of the cost function and its gradients. With -90% perturbation which is already far away from the default values of the parameters, the minimization iteratively retrieved the true parameters. The slight difference of the parameter perturbation norm is likely caused by the conditionals in the slave system. When the optimization is close to the global minimum, the adjoint model cannot provide useful gradients information to lead the optimization closer to the global minimum. Overall, the optimization is already close to the global minimum, and the parameters are accurately retrieved.

For the “minimal” configuration, the slave system can be synchronized through nudging divergence, temperature, vorticity with a nudging timescale of 2 days. A nudging timescale of 2 days is approximately the largest nudging timescale that stops the exponential error growth of the TLM. When the slave system is synchronized, the cost function becomes smooth, and the TLM reasonably describe the error propagation of the slave system. Through identical twin experiments, the adjoint model provides useful gradients of the cost function with respect to the parameters, and the parameters are accurately retrieved with an assimilation window of 1 year. When nonlinear system and observations are synchronized, the assimilation window can be extended arbitrarily.

#### **6.4 Assimilating experiments with the “maximal” configuration**

The “minimal” configuration is relatively linear with a predictability time of ~10 days. The TLM can very well describe error propagation of the slave system, and the conditional statements don’t seriously degrade the usefulness of the TLM. For the “maximal” configuration, including the moisture parameterizations such as convective precipitation, large scale precipitation, and dry convective adjustment significantly increases the nonlinearity of the nonlinear system and their implementations with multiple loops of “if” and “where” statements also degrade the usefulness of the TLM and the adjoint model. The large-scale precipitation which produces a nonzero effect when air parcels are supersaturated, is a first-order discontinuous process (Zou, 1997). This process degrades the usefulness of the adjoint model because the tangent linear approximation cannot describe the nonlinear perturbations which change the side of conditionals. The convective precipitation parameterization in PlaSim is based on Kuo (1974), which is zero-order discontinuous and the tangent linear approximation doesn’t apply theoretically. In a discrete climate model, all of these processes will lead to a discontinuous cost function. Although the exact TLM and adjoint model can be correctly generated with automatic differentiation tools, they may not be useful for the optimization.

Since 1990s’, numerical weather prediction (NWP) groups have tested different ways of dealing with moisture parameterizations in the adjoint method: (1) use the adiabatic forward model and its adjoint model (Navon et al., 1992); (2) use the full forward model to define the basic state and the cost function while using the adiabatic adjoint model for computing gradients (Zupanski, 1993); (3) use the full forward model and the simplified-physics adjoint model (Mahfouf, 1999; Rabier et al., 2000). Although the physical parameterization will make the cost function non-convex, they may not hurt the general convexity of the cost function which is assumed in gradient-based minimization algorithms (Zhang et al., 2001).

Minimization algorithms such as Quasi-Newton methods may still approach the global minima with descent directions from the adjoint model.

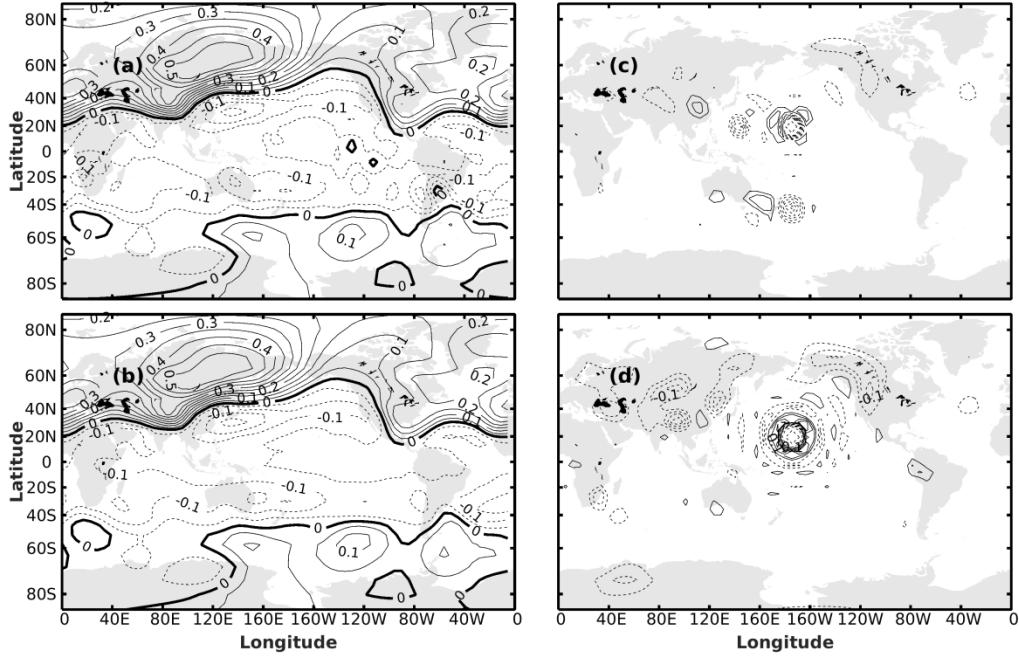


Figure 6.9. Temperature difference ( $^{\circ}\text{C}$ ) of level 01(uppermost) caused by a 10% perturbation on  $\text{tfrc1}$  after a 2-days integration of the nonlinear model (a) and the TLM (b). Temperature difference ( $^{\circ}\text{C}$ ) of level 10 (lowermost) caused by a 10% perturbation on  $\text{vdiff\_lamm}$  after a 6-hours integration of PlaSim (c) and the TLM (d).

Figure 6.9 (a) and (b) show the temperature difference for the uppermost level from perturbed PlaSim integrations and the TLM with a 10% perturbation on parameter  $\text{tfrc1}$  after a 2-day integration. The temperature differences in Figure 6.9 (a) and (b) are quiet similar for both pattern and amplitude. Fig 6.9 (c) and (d) are the same as Fig 6.9 (a) and (b) except that the perturbation is on the parameter  $\text{vdiff\_lamm}$  and the integration period is for 6 hours. The temperature differences show some similarities such as central Pacific Ocean and northwest of Northern American which indicate that the tangent linear approximation matches the nonlinear error propagation accurately without changing of the conditionals. There are considerable local changes that don't match because of switching parameterization processes in the nonlinear model on or off. But for a small time interval, we see that the TAF generated TLM can still reasonably describe error propagation of the nonlinear model before strong nonlinear moisture parameterizations are triggered. Including the nudging terms may also contribute to reduce the separation caused by these conditionals. In this chapter, the performance of the simplified DSPE method for parameter estimation will be tested with the “maximal” configuration.

#### 6.4.1 The nudging strength for synchronization

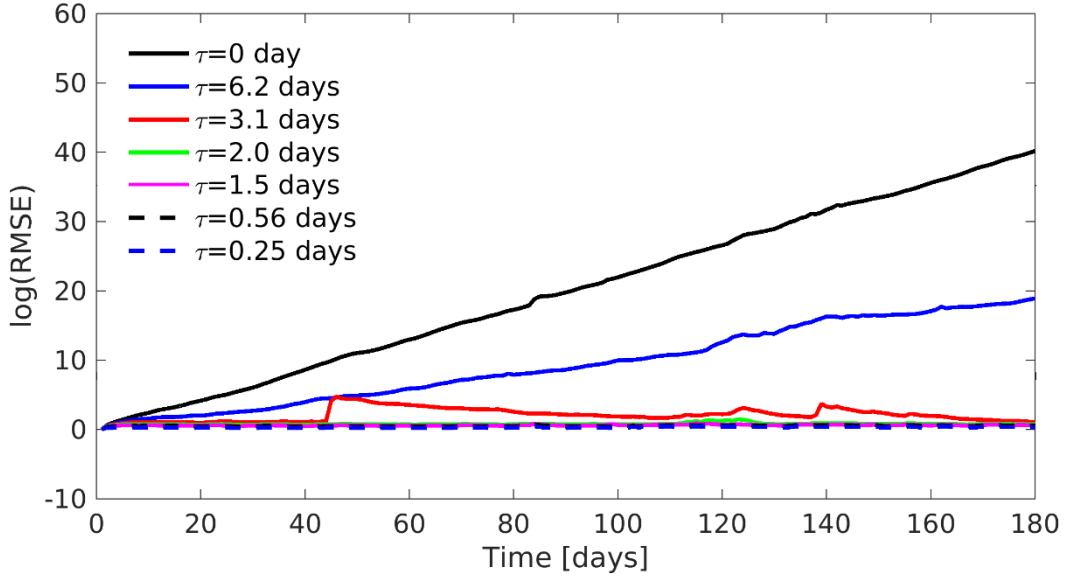


Figure 6.10. The time evolution of instantaneous RMSEs for the TLM without nudging ( $\tau=0$  day), with nudging timescales ranging from 6.2 days to 0.25 days (see legend). The “maximal” configuration is used.

As in chapter 3, it is better to predefined the best nudging coefficient for better estimating the parameters with the simplified DSPE method. But due to the high dimension of PlaSim, it is not easy to find the best nudging coefficients. For atmosphere or ocean models, usually global uniform nudging timescales are used, and the nudging timescales may depend on different purposes and different model variables. For examples, nudging timescales ranging from 0.5 days to 4 days are used to initialize decadal predictions (Polkova et al., 2014) depending on model variables. Krishnamurti et al. (1991) use nudging timescales ranging from 2.8 hours to 5.6 hours to initialize weather predictions for the tropical region. For our purpose, the nudging strength should at least be strong enough to change the positive Lyapunov exponents to negative values. In this study, a global uniform nudging timescale for all prognostic variables except surface pressure is used. The maximum nudging timescale can be decided by whether it can stop the exponential increase of the TLM error propagation. Figure 6.10 shows the logarithm of RMSEs from the TLM with a 10% perturbation on all parameters. Without nudging terms ( $\tau=0$  day), the RMSEs grow exponentially with a rate of  $\lambda=0.2550$  /days, which indicate an average predictability time of 2 days. With the increasing nudging strength, the exponential increase rate is gradually reduced. A nudging timescale of 3.1 days still shows some abrupt jumps which indicate that the synchronization is not stable. A nudging timescale of 2 days seems to be strong enough to stop the exponential error growth. In this section, the nudging strength is chosen to be two days which is a little smaller than the predictability time of the free nonlinear system.

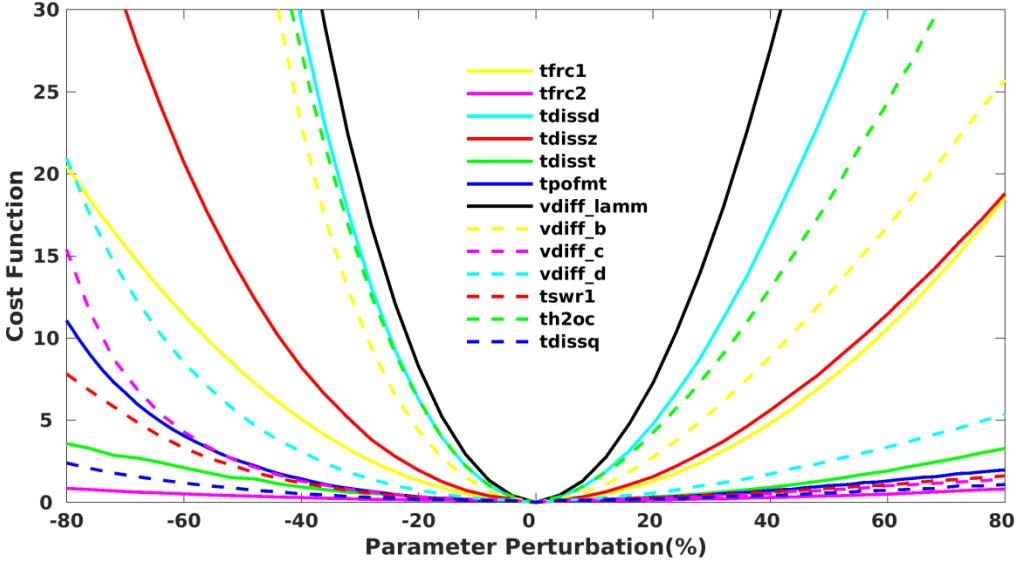


Figure 6.11. The cost functions depending on sizes of perturbations on the 13 parameters (see legend). The integration time is 2 months.

The cost functions depending on parameter perturbations for 13 parameters with a 2-month integration are shown in Figure 6.11. The relative sensitivity can be referred by the values of the cost functions divided by the size of parameter perturbation. The cost function is smooth, which indicates that the nudging strength is strong enough to prevent chaotic behavior of the nonlinear system. The parameter `vdiff_lamm` shows the strongest relative sensitivity. The parameters `th2oc`, `tdissd`, `tdissz`, `vdiff_b` and `tfrc1` also show strong relative sensitivities. For the “maximal” configuration, the cost function is not symmetric, and the left side part shows stronger relative sensitivity than the right side. For identical twin experiments, all these 13 parameters are used to test the performance of the simplified DSPE method.

In the nonlinear system, the TLM cannot well describe the error propagation when the error changes actions by the conditionals as Figure 6.9(b) shows. Although the including of nudging terms cannot fully restore the action controlled by the conditionals, they reduce the negative effects of model trajectories separation on the TLM caused by the conditionals. Before performing assimilation experiments, it is necessary to test how well the TLM can represent the nonlinear error propagation in the slave system. This is done by comparing the propagated TLM error with the finite difference of the perturbed slave system integration to the reference slave system integration. The integration time is 12 months.

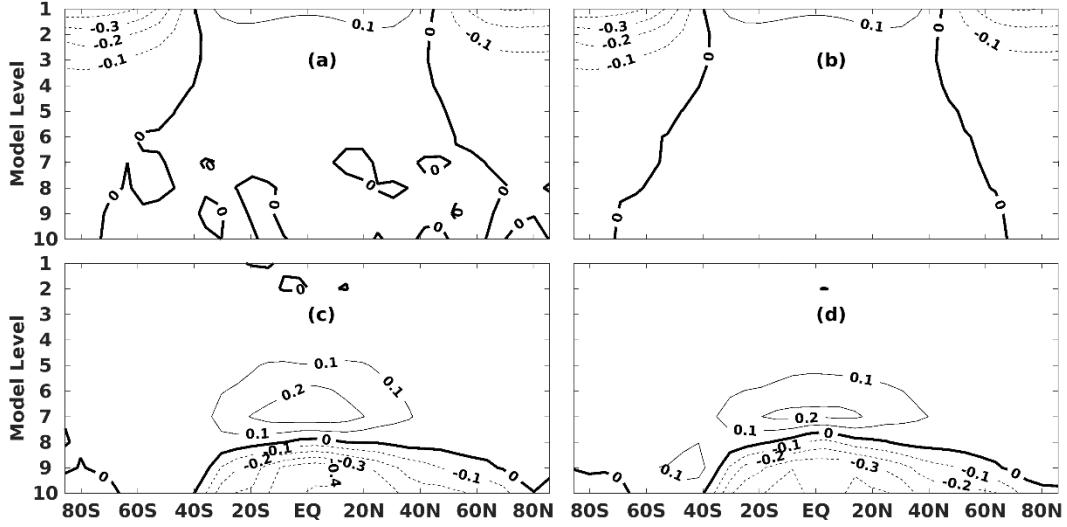


Figure 6.12. The zonal mean temperature changes from the TLM (b and d) and the finite differences of perturbed slave system integrations (a and c) with a -10% perturbation on  $tfrcl$  and  $th2oc$ : (a) and (b) are caused by perturbed  $tfrcl$ , (b) and (d) are caused by perturbed  $th2oc$ . The integration time is 1 year.

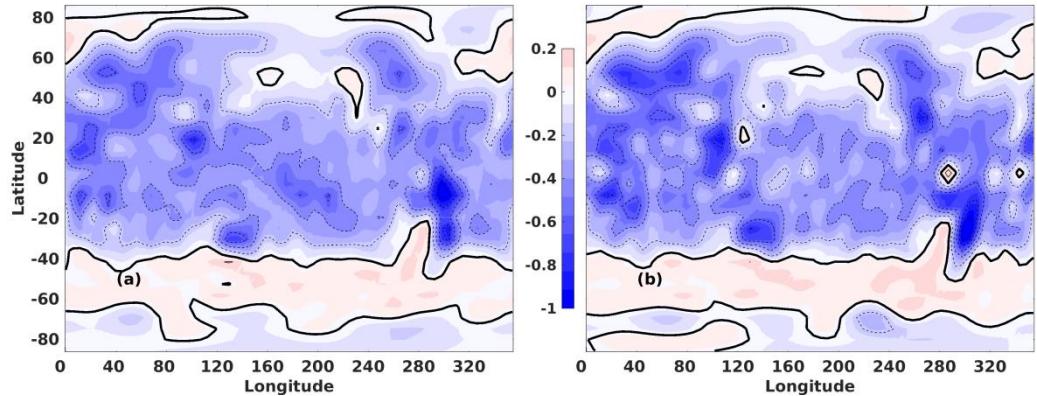


Figure 6.13. The temperature changes of level 10 for the TLM (b) and the finite difference of perturbed slave system integrations with a -10% perturbation on the parameter  $th2oc$ .

Figure 6.12 shows the zonal mean temperature change from the TLM and the finite difference caused by a -10% perturbation on the parameter  $tfrcl$  (a and b) and  $th2oc$  (c and d). The TLM represents reasonably the main characteristics of temperature change from the finite difference of the perturbed slave system integrations. There are also some different signals in the TLM and the finite difference which are likely caused by differences in on and off switching in of some parameterizations. Figure 6.13 shows the temperature differences for model level 10 with a -10% perturbation on  $th2oc$ . Both patterns and amplitude match reasonably well. Temperature is reduced for most of the region except south of 40S and north of 60N. There are also some local characteristics that don't match perfectly such as the Antarctica. But overall, the adjoint model can be expected to provide useful gradients information for the optimization. And the changes for other model states are also consistent between the TLM and the finite difference except for the specific humidity.

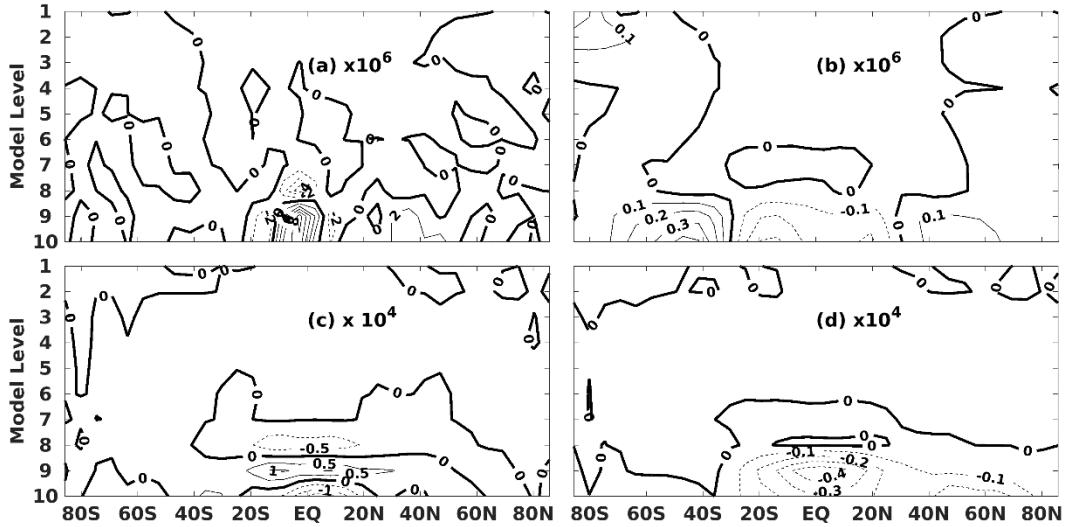


Figure 6.14. The zonal mean specific humidity (kg/kg) changes from the finite differences of perturbed slave system integrations (a and c) and the TLM (b and d) with a 10% perturbation on *tfrc1* and *th2oc*: (a) and (b) are caused by perturbed *tfrc1*, (b) and (d) are caused by perturbed *th2oc*. The integration time is 1 year.

The moisture processes are quite nonlinear and strongly influenced by the parameterization especially in the tropical region. The parameterizations are implemented with multiple loops of “if” and “where” statements which potentially degrade the usefulness of the TLM. In the tropical region, where convective precipitation dominates the variance of specific humidity, a small parameter perturbation likely changes the actions depending on conditionals. Figure 6.14 shows the specific humidity changes from the TLM and the finite difference caused by a -10% perturbation on the parameter *tfrc1* (a and b) and *th2oc* (c and d). The specific humidity changes from the TLM and the finite difference are different for both perturbed *tfrc1* and *th2oc*. The parameter *tfrc1*, which mainly influences the top of the atmosphere, has little influence on the specific humidity of lower level. Once the perturbation changes the action of the conditionals, it will significantly increase the amplitude of the changes in the finite differences as Figure 6.14 (a) and (b) show and the usefulness of the TLM and the adjoint model is seriously degraded. For the parameter *th2oc*, the TLM also provides specific humidity changes that are different from the finite difference in both pattern and amplitude.

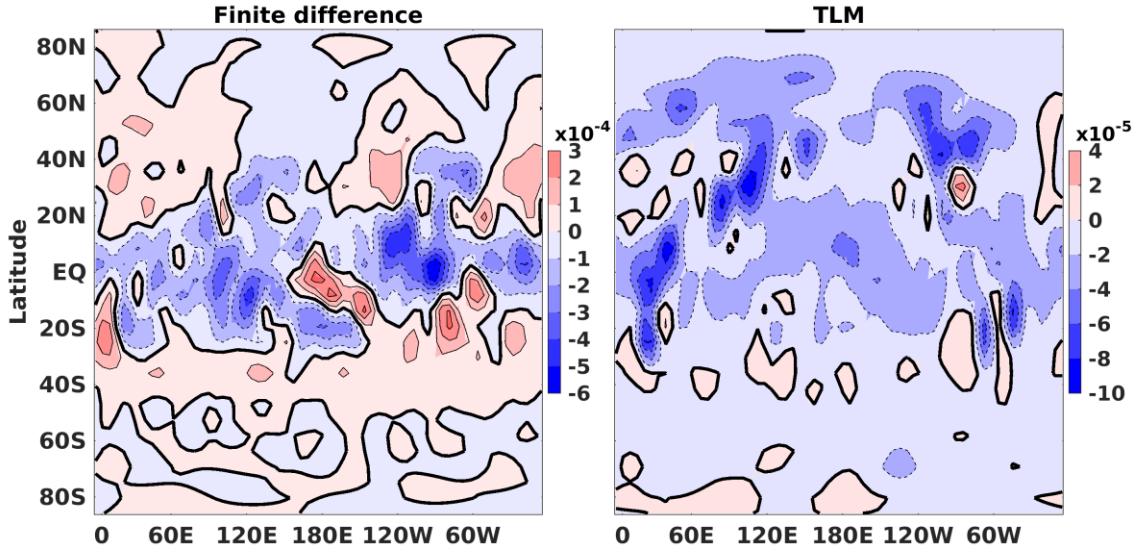


Figure 6.15. Specific humidity (kg/kg) changes in level 10 for the TLM (b) and the finite difference of perturbed slave system integrations with a -10% perturbation on the parameter  $th2oc$ . The integration time is 12 months. The scales of color bars are different for (a) and (b).

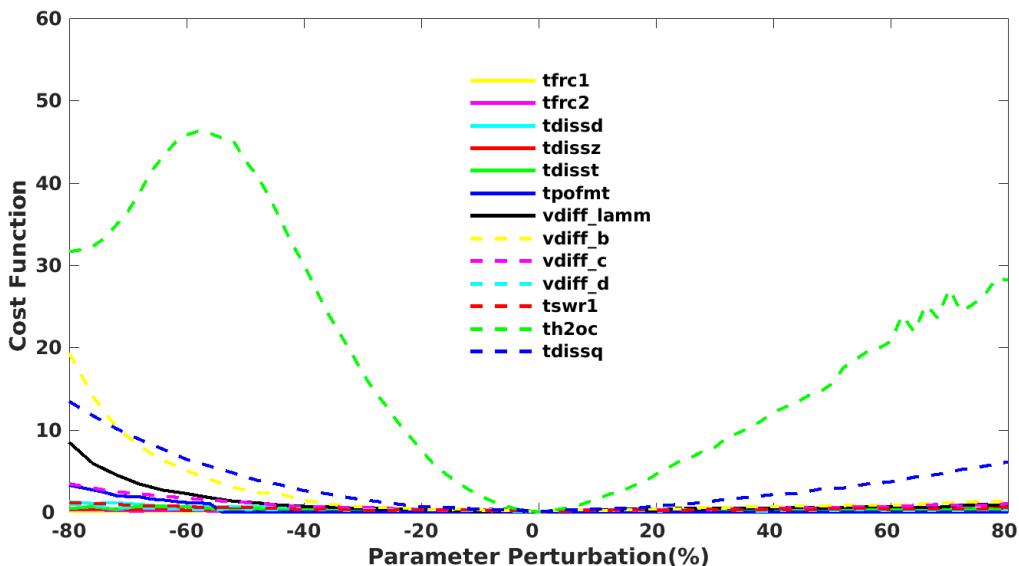


Figure 6.16. The cost function components of specific humidity depending on parameter perturbations for 13 parameters (see legend).

Figure 6.15 shows specific humidity changes for model level 10 with a -10% perturbation on the parameter  $th2oc$ . It is clear that the TLM provides specific humidity changes different with that from the finite difference of perturbed slave system integrations. The finite difference changes are of one order larger than that from the TLM, especially in the tropical region where convective precipitation dominates. Considering the consistency of the TLM and the finite difference of the slave system for other model prognostic variables, it is possible that the selected parameters cannot significantly influence specific humidity and the differences in Figure 6.15 and Figure 6.14 may just be noise. To check the sensitivity of the specific humidity with respect to the selected parameters, we show the cost functions of the

specific humidity depending on perturbations sizes for the 13 parameters are shown in Figure 6.16 with an integration time of 2 months. The cost function is relative smooth for small perturbations, but there are some zigzags for larger perturbations. In the 13 parameters, there are two parameters that show considerable relative sensitivity with respect to the specific humidity:  $tdissq$  and  $th2oc$ . But as Figure 6.14 and Figure 6.15 show, the TLM cannot reasonably describe the impacts of the parameters on the specific humidity. However, the TLM can still accurately describes the impacts on other model states as Figure 6.12 and Figure 6.13 show. Hence, the specific humidity is not included in the cost function, but it is still used to synchronize chaos of the nonlinear model and observations.

#### 6.4.2 Assimilation experiments

As shown above, the simple nudging terms can successfully synchronize chaos with nudging timescales smaller than two days by nudging temperature, vorticity, specific humidity and divergence. The TLM can also reasonable represents the error propagation of the slave system for all model states except for the specific humidity. In this section, the performance of the simplified DSPE method is tested through identical twin experiments. Four experiments are performed with perturbation sizes: 10%, -10%, 20%, -20% on all 13 parameters. The assimilation window is two months. The successfulness of the assimilating experiments is judged by whether the default values of the parameters are accurately retrieved.

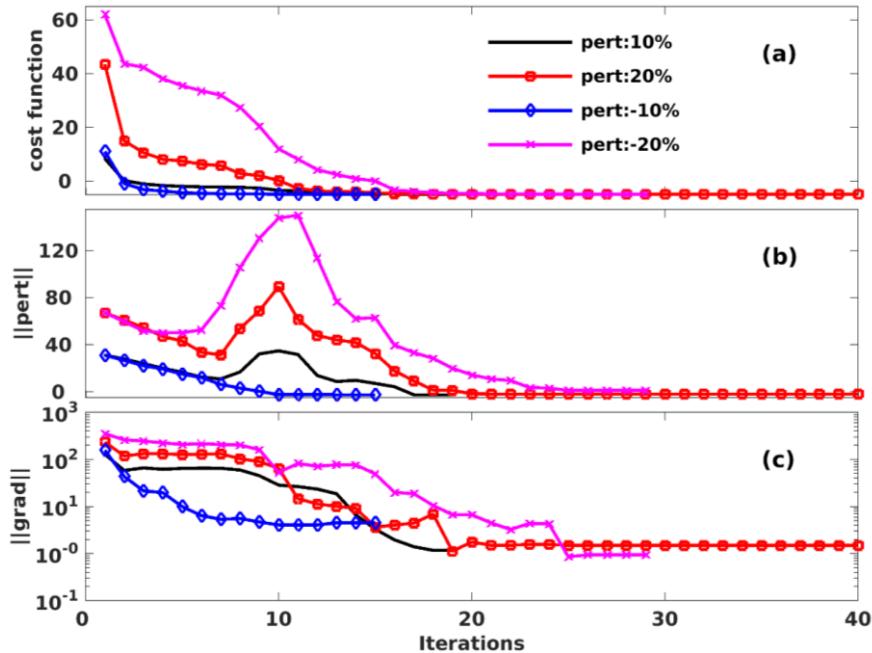


Figure 6.17. Convergence of the minimization for 4 experiments with perturbation sizes of 10%, -10%, 20% and -20% (see legend): (a) the cost function, (b) norm of the perturbation sizes, (c) norm of the gradients.

Figure 6.17 shows the convergence of the minimization for all four experiments. As Figure 6.17 (a) and (c) show, the cost functions for all four experiments are reduced close to 0, and their gradients are also significantly reduced after the optimization. The optimization is

stopped when the adjoint model cannot provide useful gradients, and the final gradients are slightly different as Figure 6.17(c) shows. But as indicated by the cost function and absolute differences of the parameters to their true values (Figure 6.17 (b)), the four assimilating experiments accurately retrieves the default values of the parameters, and the minimization is likely to close to the global minimum at last. The assimilating experiment with a -10% perturbation on all parameters is an ideal case in which both the parameter norm and the gradients are iteratively reduced. In experiments with perturbation sizes of 10%, 20% and -20%, the absolute differences of the control parameters to their true values are firstly decreased for initial iterations and then increased from the 6th iteration. At last, the absolute differences are reduced close to 0.

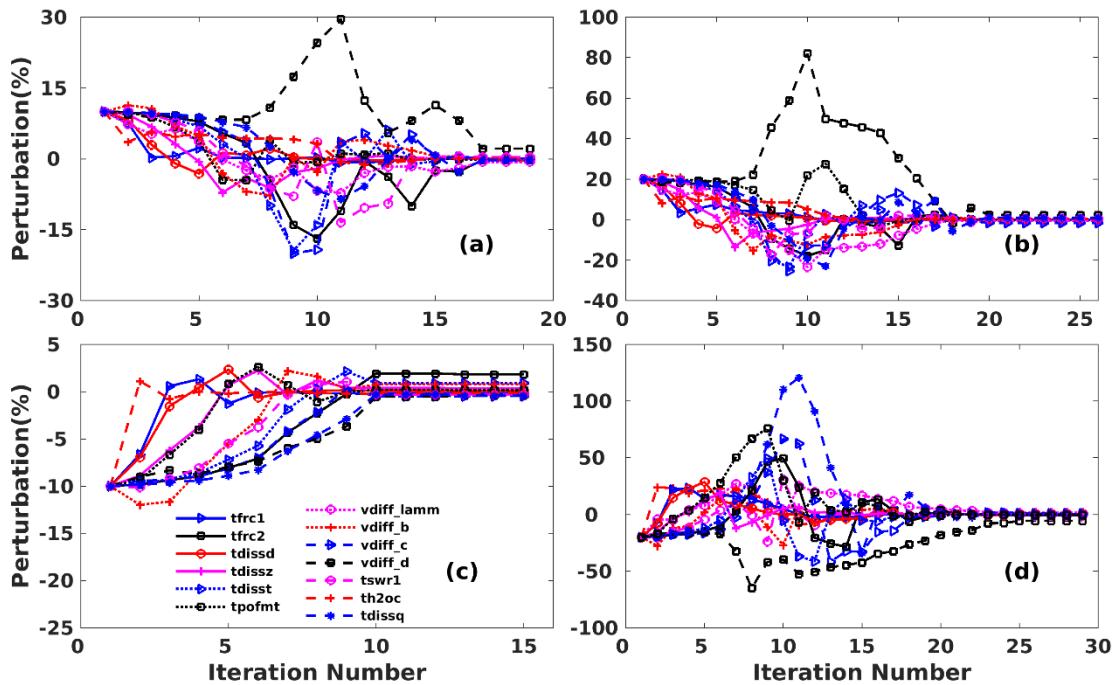


Figure 6.18. Absolute difference of components of control parameters over iteration number.

The norm of perturbations for all parameters and the gradients with respect to each of the parameter are plotted in Figure 6.18 and Figure 6.19. For the assimilating experiment with a -10% perturbation on all parameters as shown in Figure 6.18(c) and Figure 6.19(c), the minimization gradually brings the parameter perturbations close to 0. After ten iterations, the absolute difference of the estimated parameters to the true values cannot be further reduced, and the optimization is finally stopped after the 15<sup>th</sup> iteration. The minimization cannot further approach the global minimum because the adjoint model cannot provide useful gradients. For the other three experiments, the parameter perturbations are gradually reduced during initial iterations. Then, the absolute differences of several parameters such as *vdiff\_d*, *vdiff\_c*, *tdisst*, *tdissq* to the true values are significantly increased, and the norm of the perturbation sizes are increased as Figure 6.17(b) shows. As Figure 6.19, these parameters have small sensitivities. While for parameters with larger sensitivities such as *vdiff\_lamm*, *th2oc*, *tfrcl* and *tdissz*, the optimization quickly reduces the absolute difference of these parameters to their true values.

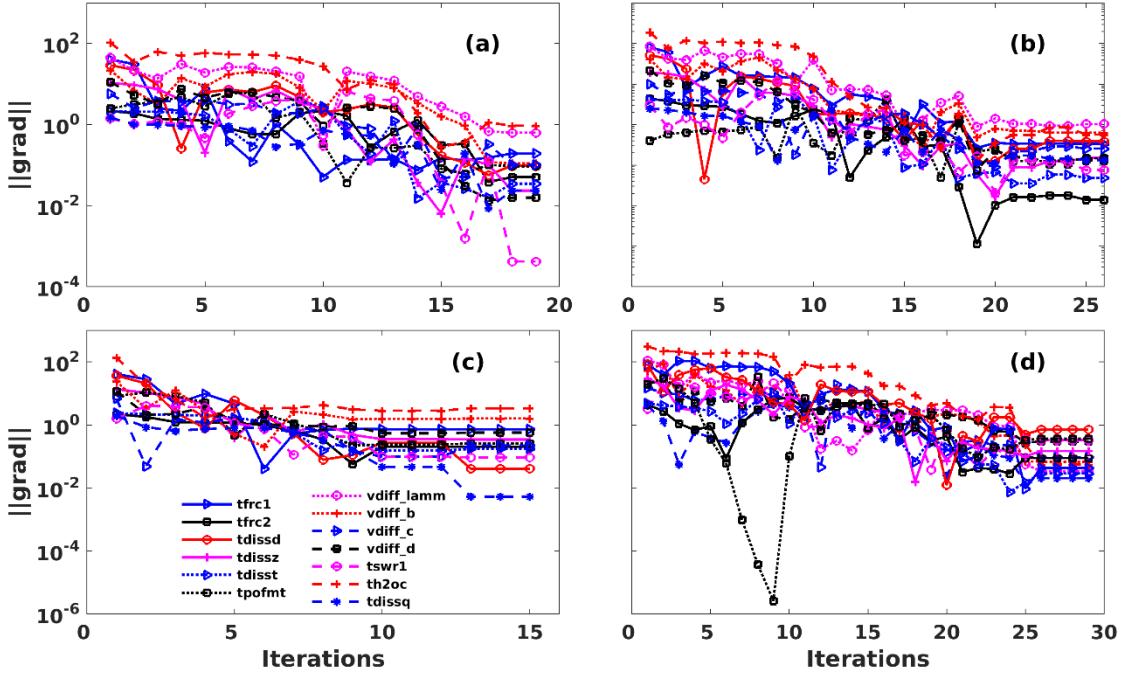


Figure 6.19. As Figure 6.18 but for gradients of the cost function with respect to the control parameters.

In the Quasi-Newton algorithm used in this study, the minimization directions are decided as equations (3.20)-(3.23) with the gradients provided by the adjoint model. Here, the parameter  $tdissd$  in the experiments with -20% perturbations which is increased significantly during the optimization is taken as an example to show the performance of the minimization algorithms. For computing the new searching direction for the minimization, the new approximated inverse of the Hessian matrix and the gradients from the adjoint model are used as in equation (3.23). The new gradient, the gradient from the adjoint model, the diagonal part and the off-diagonal part for the parameter  $tdissq$  are shown in Figure 20. As Figure 6.19(d) shows, the gradient of the cost function with respect to the parameter  $tdissq$  is much smaller than most of the parameters. During iteration 2-10, the off-diagonal part decides the new searching direction and provides a negative value of the gradient for this parameter. Hence, the parameter  $tdissq$  is increased during these iterations but has little effects on the cost function due to its small sensitivity. In the 8<sup>th</sup> iteration, the gradient shows abrupt reduction which is likely due to the nonlinear and discontinuous parameterizations. But at last, the Quasi-Newton algorithm successfully reduces the parameter perturbations to 0 with the gradients information provided by the adjoint model.

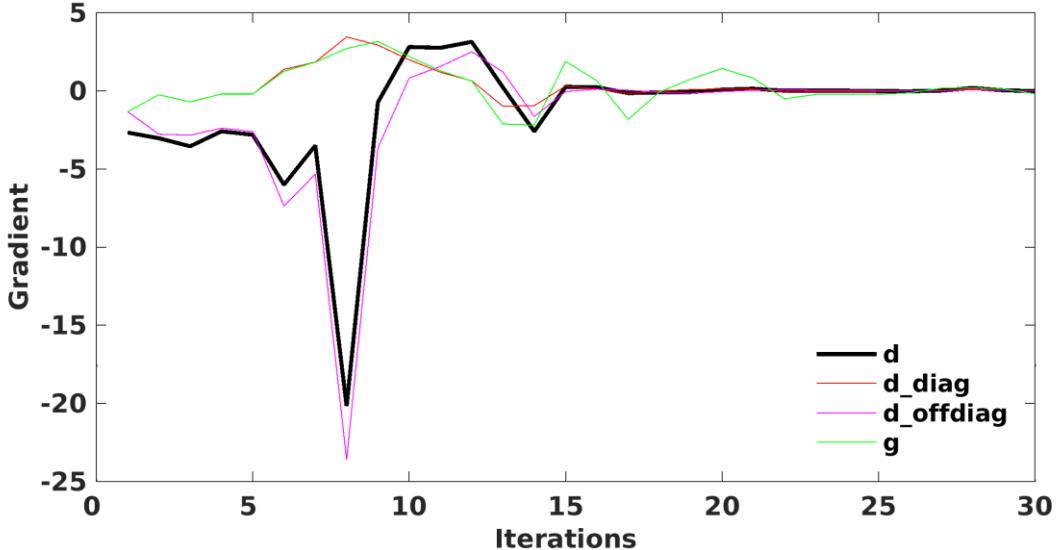


Figure 6.20. The new search directions ( $d$ ), the diagonal part ( $d_{\text{diag}}$ ) and off-diagonal ( $d_{\text{offdiag}}$ ) and the gradients information from the adjoint model ( $g$ ) for the parameter *tdissq.in* experiments with -20% perturbations. The components are computed based on equations (3.20)-(3.23).

For the “maximal” configuration, although the strong nonlinearity and conditionals in the moisture parameterizations, the adjoint model still provides usefulness gradient information for the minimization. Figure 6.11 shows that the cost functions for some parameters are not quadratic, but the assimilating experiment results show that the minimization algorithm which assumes quadratic cost function can still iteratively bring the parameters back to their default value. The optimization will finally exit when the parameter errors are very small. At that time, the adjoint model cannot provide useful gradients information which can lead the minimization further approaching the global minimum. For all the four identical twin experiments, the default values of the parameters are already reasonably retrieved. Overall, the TAF generated adjoint model can provide useful gradients, and the simplified DSPE method works for the “maximal” configuration.

#### 6.4 Discussions and conclusions

In this part, the simplified DSPE method is tested with the more complex chaotic system PlaSim. Synchronization is achieved by simple nudging (Hoke and Anthes, 1976) technique. Two configurations are used to test the performance of the simplified DSPE method. One is the “minimal” configuration in which moisture parameterizations are excluded, and the climatological surface temperature is used to replace the slab ocean and the land model, respectively. The other one is the “maximal” configuration, which includes all the moisture parameterizations and is, therefore, more nonlinear. With strong enough nudging terms, the feasible assimilation window can be extended arbitrarily, and the parameters are accurately be retrieved for both the “minimal” configuration and the “maximal” configuration in the identical twin experiments. Due to the strong nonlinearity of vertical diffusion parameterizations, directly linearizing these processes will degrade the usefulness of the adjoint model, even lead to the adjoint model unstable, and modifications are made to

improve the TLM and its adjoint (Errico et al., 1993; Laroche et al., 2002; Mahfouf, 1999; Zhu and Kamachi, 2000; Zhu et al., 2002).

For the “minimal” configuration, the predictability time is around 10 days which allows feasible assimilation windows around 20 days. By comparing the nonlinear error growth and the tangent linear error growth, a potential problem in the TAF generated TLM model associated with the linearization of the vertical diffusion parameterization is found. As suggested by Zhu et al. (2002), this problem is solved by setting  $R'_i = 0$  in the TLM and its adjoint. By using a nudging timescale of 2.0 days, the chaotic behavior of the nonlinear system and the pseudo-observations is synchronized. The cost function is smooth but not absolute symmetric due to the nonlinearity of the slave system. The TLM accurately describes the nonlinear error propagation of the slave system. Identical twin experiments with an assimilation window of 1-year show that the modified DSPE method works perfectly despite some of the conditionals in some of the parameterizations.

For the “maximal” configuration, the inclusion of the moisture parameterizations increases the nonlinearity of the system. The predictability time is reduced to around 4 days which allows feasible assimilation windows of 8 days. Blessing et al. (2014) have shown that the feasible assimilation window is limited to no more than 1 day. The reason is that the moisture parameterizations are implemented with multiple loops of “if” and “where” statements, which degrade the usefulness of the TLM and its adjoint. But for finite perturbations, the TAF generated TLM can still describe nonlinear error propagation if the conditionals are not changing any action due to a perturbation. With a nudging timescale of 2.0 days and nudging temperature, vorticity, divergence and specific humidity, the system is synchronized which is identified from the non-exponential error growth of the TLM. By comparing the error growth of the TLM and the finite difference of the perturbed slave system integrations, it is shown that the TLM can still accurately describe the nonlinear error propagation of temperature, divergence, vorticity and surface pressure except for the specific humidity due to the multiple loops of “if” and “where” statements. By using temperature, vorticity, divergence and surface pressure in the cost function, the parameters are successfully retrieved for a 2-month assimilation window

Overall, the simplified DSPE method still works for this complex climate model PlaSim. Although there are multiples loops of “on-off” switches in moisture parameterizations, the TAF generated adjoint model can provide useful gradients information for the optimization, and the parameters are reasonably well recovered.

# Chapter 7

## Assimilating the ERA-Interim data

As the identical twin experiments above show, we retrieved the default values of the parameters with accurate and high-frequency pseudo-observations, using the modified DSPE method. Compared to the ERA-Interim reanalysis data, our model resolution is coarser, and the parameterizations are simpler. Assimilating the ERA-Interim data is much more challenging due to the unknown model errors, observation noise, and interpolation used to fill the 6-hour gaps between two reanalysis times. The estimated parameters may be influenced by the finite nudging terms. For parameter estimation on climate timescales, it is likely that the tuned parameters compensate error from other parameterizations. Therefore, with a different combination of control parameters and with different observations in the cost function, the parameters may not converge to the same values after the optimization. In this section, the modified DSPE method will now be applied to assimilate the ERA-Interim reanalysis dataset (Dee et al., 2011). First, the slave system with nudging terms will be used to estimate the control parameters by assimilating the ERA-Interim data. Then, contributions of each parameter to the cost function reduction will be analyzed carefully to check how the method works. At last, effects of the estimated parameters on the free system (without nudging terms) must be evaluated.

The questions to be investigated are: (1) by using all the prognostic variables as in the identical twin experiments, can we reasonably estimate the model parameters? (2) By including longwave (LW) and shortwave (SW) fluxes at the surface and the top of the atmosphere (TOA) in the cost function, can the method reduce the flux biases by tuning clouds properties, and at the same time, improves the model-simulated temperature. (3) Can the estimated parameters improve the free model's climatology? This part will provide a guideline for further applying this method to the coupled model CESAM.

### 7.1 Experiment setup

Due to the complex model and the noisy observations, it is impossible to define the best nudging timescale as Figure 4.5 displays. But the nudging timescales should be smaller than 2.03 days in the ‘maximal’ configuration as Figure 6.10 shows. Because of the model error, the nudging terms may be larger than the model’s physics terms. For a nudging timescale of  $\tau=2.03\text{days}$ , the ratios of the nudging terms to the total tendencies for temperature (a), moisture (b), divergence (c) and vorticity (d) are shown in Figure 7.1. The nudging terms are large in all four variables, especially in the tropics and in the polar region. With smaller nudging timescales, the ratios will be larger which may not benefit the optimization. The large ratios also indicate that the model has large errors which may not be absolutely reduced by optimizing the process parameters. Therefore, the effects of the estimated parameters on the free model (without nudging terms) must be evaluated after the optimization.

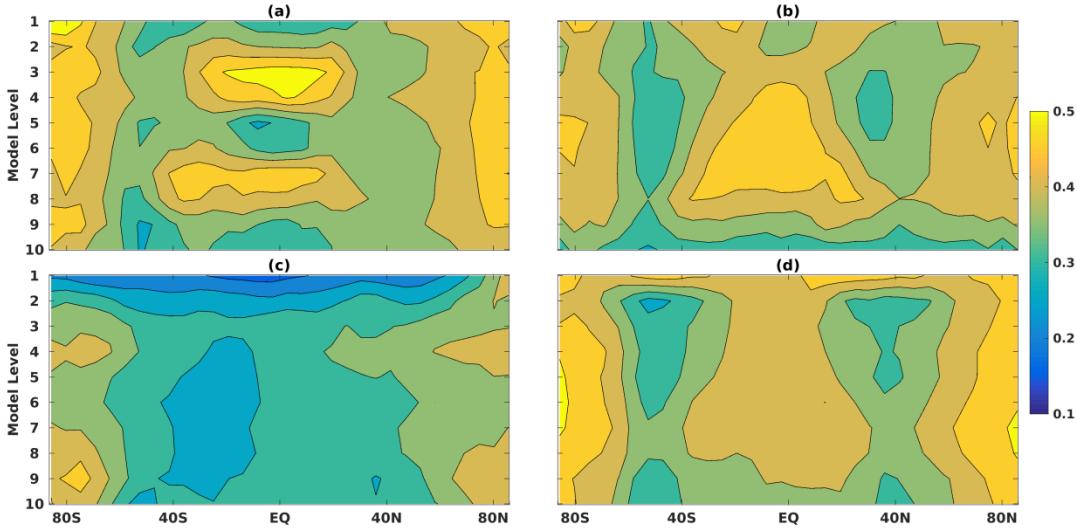


Figure 7.1. Ratios of the nudging terms to the total tendencies for temperature (a), moisture (b), divergence (c) and vorticity (d). The ratios are averaged over a 6-month integration period.

In this chapter, two groups of assimilation experiments are performed. The first group assimilation experiments consist of 6 experiments with nudging timescales of 2.03 days, 1.19 days, 0.56 days, 0.25 days, 0.15 days, and 0.09 days. A nudging timescale of 2.03 days is a little larger than the leading Lyapunov exponent as Figure 6.10 shows and is of the same order of damping timescales, while a nudging timescale of 0.09 days is very strong and is of the same order as for initializing the NWP in the tropical region (Krishnamurti et al., 1991). Similar to the identical twin experiments in chapter 6, temperature, divergence, vorticity and specific humidity from the ERA-Interim data are used in the nudging terms to push the PlaSim simulation to follow the ERA-Interim data while annually averaged temperature, divergence, vorticity and surface pressure are used in the cost function. The 16 parameters listed in Table 5.1 are chosen as the control parameters. The performance of the optimization and influences of the nudging timescales are investigated based on these assimilation experiments.

Based on results of the first group assimilation experiments, the second group assimilation experiments consisting of two experiments are performed. The two assimilation experiments use the same variables as in the first group experiments for synchronization while they use different control parameters and different observations to compute the cost function. Details of the two experiments are listed in Table 7.1. Expt-p04 includes four control parameters: *tfrc2*, *vdiff\_d*, *th2oc* and *acllwr* and uses net LW flux at the surface and at the TOA and air temperature to compute the cost function. Parameters *tfrc2* and *vdiff\_d* are selected due to their effects on temperature based on results of the first group assimilation experiments. Parameters *th2oc* and *acllwr* tune LW absorption coefficient of water vapor and clouds and will be optimized to reduce the large biases of net LW fluxes at the surface and at the TOA. Expt-p07 further takes parameters *tswr1*, *tswr2* and *tswr3* as control parameters to reduce net SW flux biases.

Table 7.1 Experiment setups for the second groups of assimilation experiments.

	<b>Control parameters</b>	<b>Observations in cost function</b>
<b>Expt-p04</b>	<i>tfrc2,vdiff_d,th2oc,acllwr</i>	Top LW flux; Surface LW flux; Air temperature
<b>Expt-p07</b>	<i>tfrc2,vdiff_d,th2oc,acllwr,tswr1,tswr2,tswr3</i>	Top LW flux; Surface LW flux; top SW flux; surface SW flux; Air temperature

The ERA-Interim data is first vertically averaged to our model grid and then horizontal spectral truncated to the T21 grid for nudging and computing the cost function. The PlaSim is spun up for 1 year with active nudging terms to push the model closer to the observations, and the restart file is stored. Afterwards, the assimilation experiments are performed starting from this restart file, and the assimilation window is 1 year. When the nudging terms are added to the model, the relative importance of the cost function components is also distorted. To make full use of all observations during the optimization, weighting factors are added to different components of the cost function and make them equally-weighted which leads to an initial value of the cost function equal to 15.5.

## 7.2 Results for the first group assimilation experiments

### 7.2.1 Convergence of the minimization

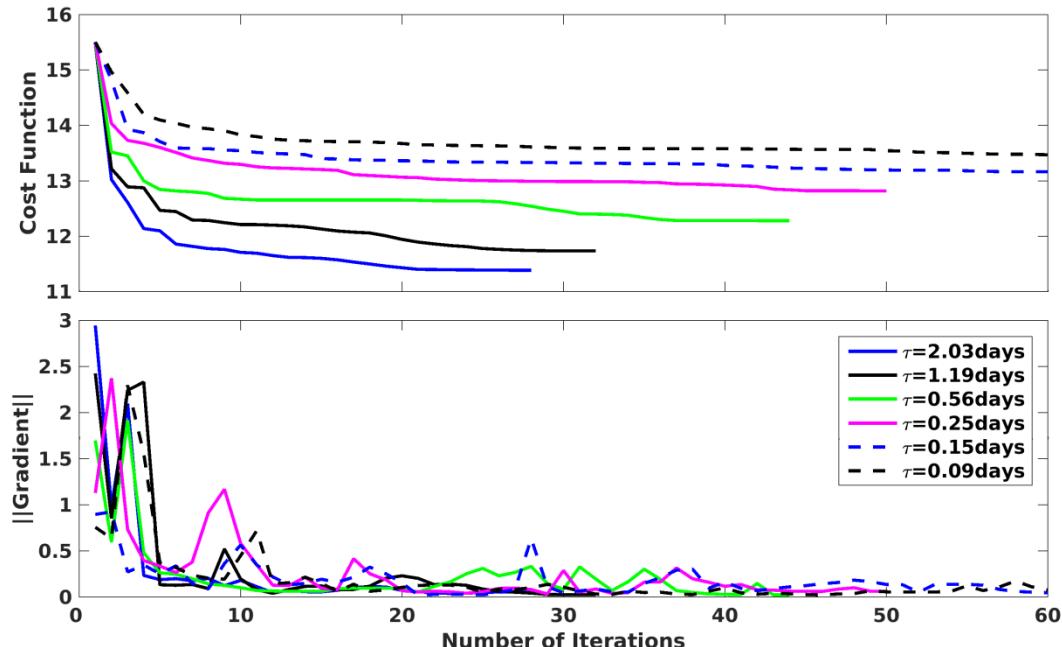


Figure 7.2. Convergence of the minimization for the cost function and the gradient norm in all six experiments.

Figure 7.2 shows the convergence of the minimization in all six experiments with different nudging timescales (see legend). With a nudging timescale of 2.03 days, the optimization quickly reduced the cost function from 15.5 to 11.4 within 29 iterations. The minimization stopped because the adjoint model cannot provide useful gradients. There are

still considerable errors that cannot be reduced. With decreasing nudging timescales, the reduction of the cost function is reduced, and iteration numbers are increased. Because with decreasing nudging timescales, the sensitivities of the cost function with respect to the parameters are reduced. The gradient norms are also iteratively reduced, and there are some abrupt increases due to nonlinearity. From the perspective of the cost functions and gradient norms, the minimization helps to reduce the difference between the slave system and the ERA-Interim data.

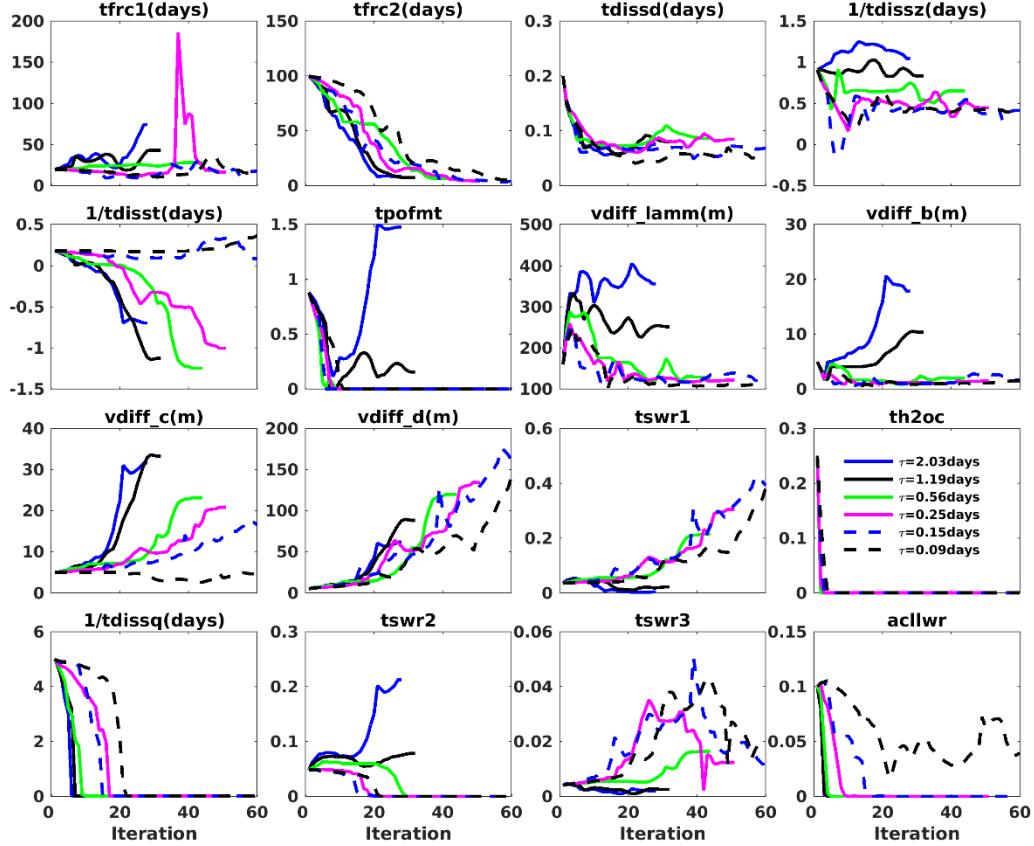


Figure 7.3. Convergence of all 16 parameters (see title) in the 6 experiments.

Convergence of the 16 parameters in the six experiments is displayed in Figure 7.3. Except for parameters  $tfrc2$ ,  $tdissd$ ,  $th2oc$  and  $tdissq$  which converge to similar values respectively in all six experiments, the estimated values of the other parameters show divergence and seems to depend on the nudging timescales. Parameter  $th2oc$  approached 0 after the first iteration for all experiments which means that the slave system doesn't need this extra exponential absorptivity of long wave radiation by water vapor. Although  $tdissq$  drifts to infinite ( $1/tdissq \sim 0$  days) in the six experiments, which indicates no diffusion for specific humidity, this is not physically meaningful. For the rest of the parameters, we see transitions with decreasing nudging timescales. For example, the estimated values of parameters  $vdiff\_lamm$ ,  $vdiff\_b$  and  $vdiff\_c$  are gradually reduced with decreasing nudging timescales. Parameter  $tdisst$  is finally tuned to unphysical values ( $tdisst > 0$ ). Parameters  $tswr2$ ,  $tswr3$ , and

*acllwr* in some experiments are tuned to 0 which is also out of their range and may have significant impacts on the radiation processes.

Overall, all the six experiments successfully reduce the cost functions and the gradient norms. Among all the 16 parameters, parameters *th2oc*, *tfr2c2*, and *tdissd* converge to similar values in all the six experiments while the other parameters show divergence or are even be tuned out of their physical range. For the minimization, the cost function reduction decreases with decreasing nudging timescales because the sensitivities of parameters are reduced.

### 7.2.2 Evaluation of the optimization

To better illustrate the effects of the parameter estimation on different variables of the slave system, the temperature difference between the slave system and the ERA-Interim data and the temperature increment after the optimization are shown. Here, we only show the results of the assimilation experiment with a nudging timescale of  $\tau=2.03$  days. For the air temperature, the other experiments show similar results. Figure 7.6 (a) and (b) show zonally averaged temperature difference between the slave system and the ERA-Interim data and the temperature increment after the optimization. The slave system has a warm bias in the tropical region centered at level 03, and in the polar region for the uppermost layer. Except for level 10 and level 9, most of the regions are still colder than the ERA-Interim data. As Figure 7.6 (a) shows, the warm bias of the slave system in level 10 and level 9 are mainly over land regions which also exists in the free model as shown in Figure 5.3. By using simple nudging terms to couple PlaSim and the ERA-Interim data, the model-observation misfits over land regions are harder to reduce than over ocean regions, especially in the northern hemisphere in level 09 and level 10.

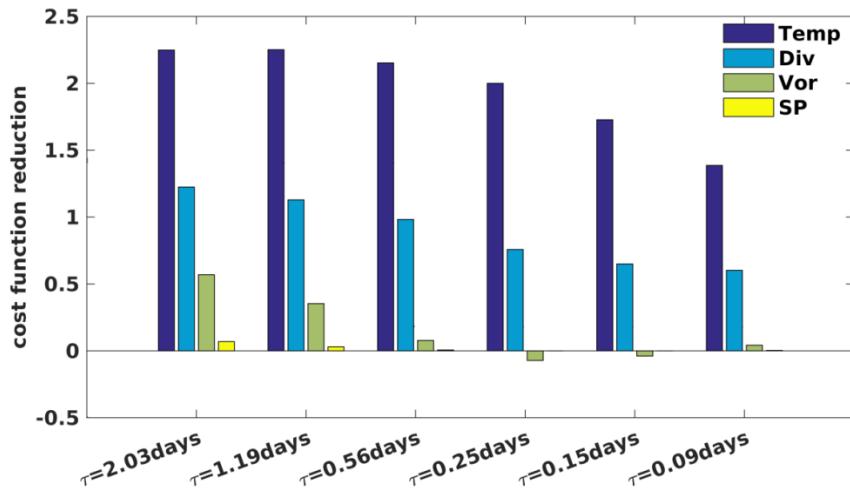


Figure 7.4. Constituents of the cost function reduction in the six assimilation experiments.

The temperature increments after the optimization are shown in Figure 7.5 (a) and Figure 7.6 (b). The misfits over the land regions in level 10 and level 9 are reduced by 2 °C and the cold bias in middle levels is reduced by around 1.0 °C. The warm bias centered in level 3 in the tropical is slightly reduced by 0.5 °C. The Antarctica shows both improvement and degradation after the optimization. In other five experiments with different nudging

timescales, the temperature differences and the temperature increments show similar patterns with that shown in Figure 7.5 (a) and Figure 7.6 (a), especially in the tropics and over the land regions, but the amplitudes are reduced with decreasing nudging timescales. However, the differences and the increments of divergence, vorticity and surface pressure varies a lot, especially the patterns, with decreasing nudging timescales.

### *a. The air temperature*

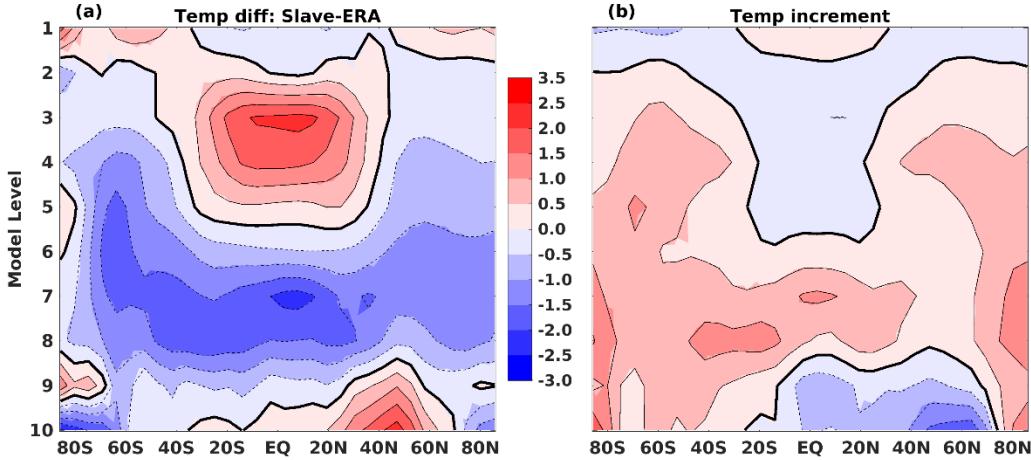


Figure 7.5. (a) Zonally averaged temperature difference ( $^{\circ}\text{C}$ ) of the slave system simulation and the ERA-Interim data. (b) Zonally averaged temperature increment after the optimization in the assimilation experiment with a nudging timescale of  $\tau=2.03$  days. The contour interval is  $0.5\text{ }^{\circ}\text{C}$ .

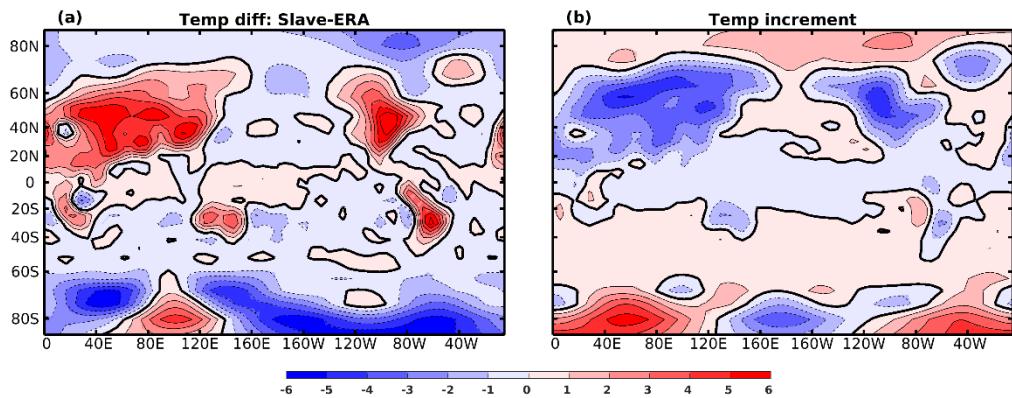


Figure 7.6. (a) The temperature difference ( $^{\circ}\text{C}$ ) in level 10 between the slave system simulation and the ERA-Interim data. (b) The temperature increment ( $^{\circ}\text{C}$ ) in level 10 after the optimization of the slave system. The nudging timescale of the slave system is  $\tau=2.03$  days. The contour interval is  $1\text{ }^{\circ}\text{C}$ .

### *b. The other variables*

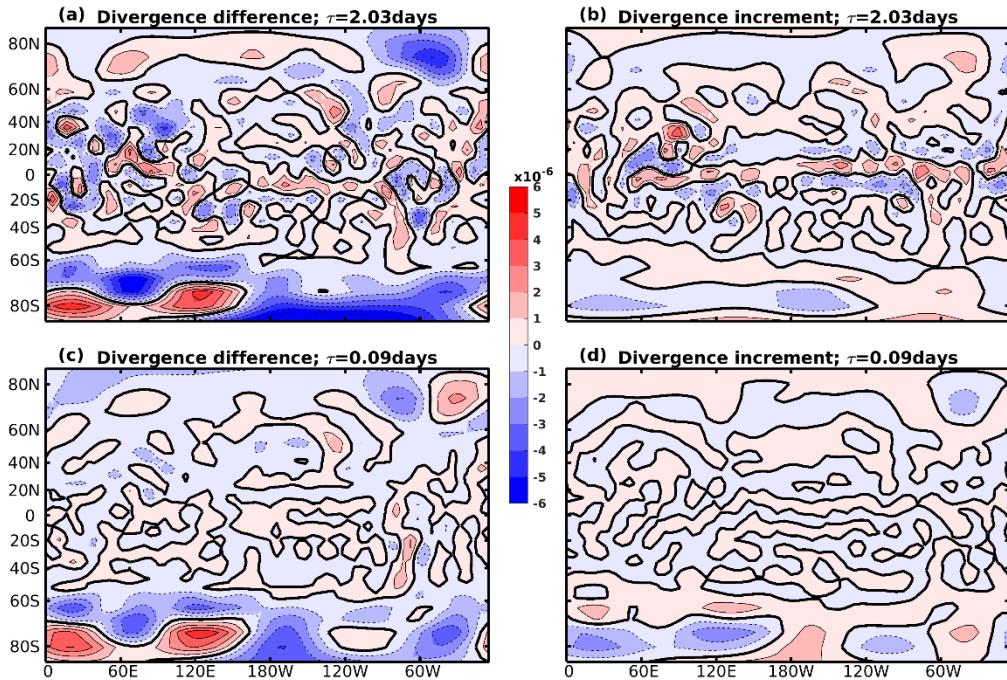


Figure 7.7. The divergence differences between the slave systems and the ERA-Interim data with nudging timescales of  $\tau=2.03\text{days}$  (a) and  $\tau=0.09\text{days}$  (c) in model level 10. The divergence increments after the optimization are shown in (b) and (d), respectively.

Figure 7.7 shows the divergence differences and the increments in the assimilation experiments with nudging timescales of  $\tau=2.03\text{days}$  and  $\tau=0.09\text{days}$ . With decreasing nudging timescales, the divergence difference is reduced especially in the tropics while still considerable large south of 60S. In the assimilation experiments with a nudging timescale of  $\tau=2.03\text{days}$ , the optimization reduces the biases in the tropics and the Antarctica. While in the assimilation experiments with a nudging timescale of  $\tau=0.09\text{days}$ , the optimization only reduces the error in the Antarctica with little effects in the tropics. Because in the slave system with a nudging timescale of  $\tau=0.09\text{days}$  the bias information is damped out by the strong nudging terms and the sensitivities of divergence observations with respect to process parameters are significantly reduced. Compared with the air temperature observations, the divergence observation error information seems more easily be damped by the nudging terms.

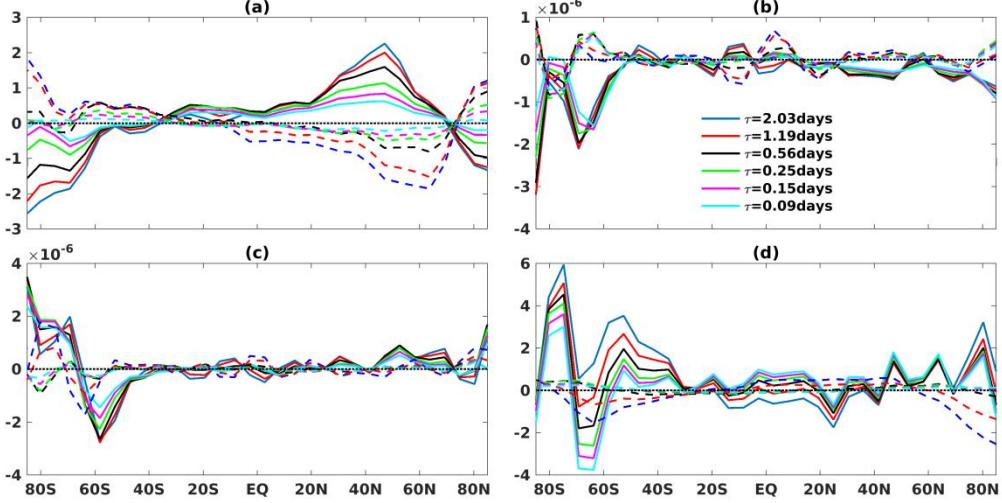


Figure 7.8. Zonally averaged differences and zonally averaged increments of temperature (a), divergence (b), vorticity (c) and surface pressure (d) in model level 10. The solid lines represent the differences of the slave system with different nudging timescales and the dashed lines are the increments, respectively. The colors represent the assimilating experiments with different nudging timescales as the legend indicates.

Figure 7.8 shows zonally averaged differences and zonally averaged increments of temperature (level 10), divergence (level 10), vorticity (level 10) and surface pressure in all the six assimilation experiments. The slave systems have a cold bias in the polar region and a warm bias in the northern hemisphere, especially over the land regions as Figure 7.6 shows. The optimization tries to reduce the bias with temperature increments that are opposite to the temperature differences. With decreasing nudging timescales, the temperature differences are reduced which, at the same time, reduce the sensitivities of parameters and thus the temperature increments after the optimization are also reduced. Overall, the optimization reduces the temperature bias of the slave systems by tuning the parameters.

As Figure 7.7(b) shows, the divergence difference is reduced south of 60S and in the tropics. With decreasing nudging timescales, the divergence difference south of 60S is hardly reduced while are quickly reduced in the tropics and even invisible with a nudging timescale  $\tau < 0.56$  days. As Figure 7.7(b) and Figure 7.8(b) display, the divergence differences in the tropics can be reduced by the optimization while they are only slightly reduced south of 60S. The large divergence difference south of 60S is likely related to the coarse model resolution and is hardly removed by only tuning process parameters. The vorticity differences and the surface pressure differences as shown in lower of Figure 7.8 are also hardly reduced after the optimization. After synchronization, it seems that the divergence observations, the vorticity observations and the surface pressure observations are much too noisy to be used for the optimization. Also, the errors seem to related to model deficiency and cannot be reduced by optimizing the process parameters.

### 7.2.3 Detailed effects of each parameter

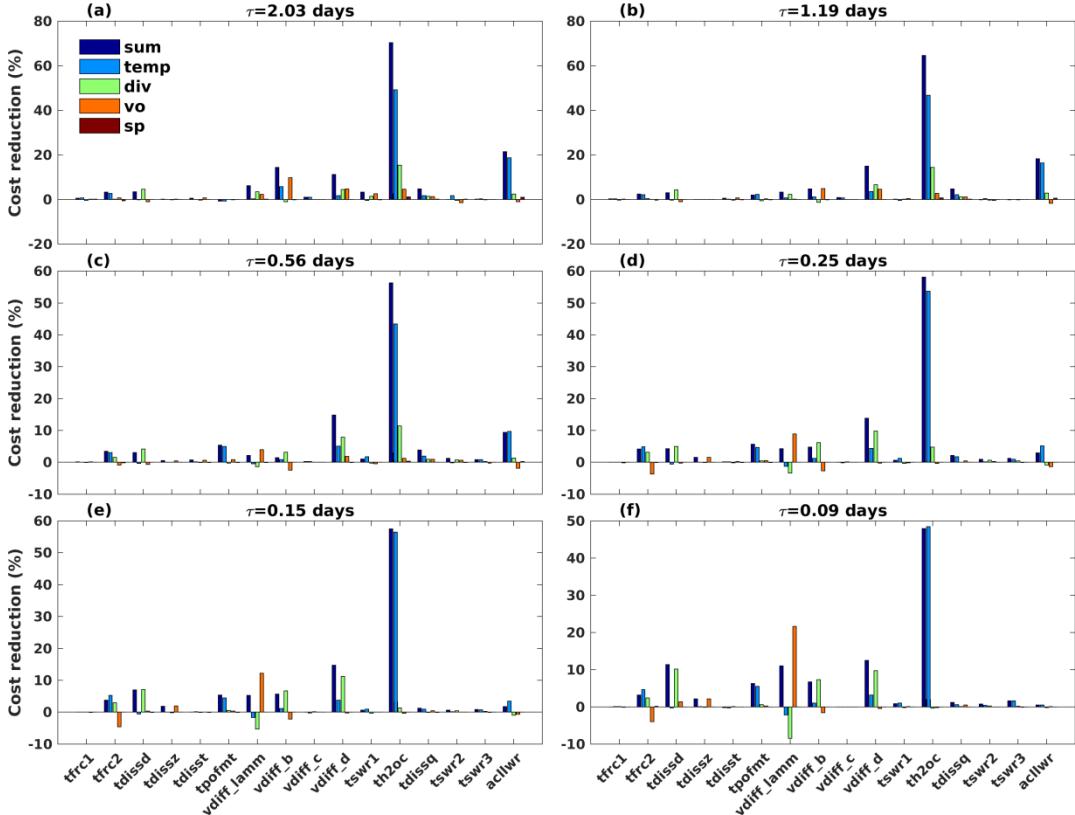


Figure 7.9. Contribution of each parameter (see legend) to constituents of the total cost function reduction in all 6 experiments (see titles).

To further understand the contribution of each parameter to the slave system, the 16 parameters are perturbed from their optimized value to their default values one by one for all the six experiments. The contribution of each parameter to the reduction of the cost function is approximated by the cost function difference between each perturbed integration and the optimized integration divided by the cost function reduction, respectively. Due to the nonlinearity of the slave system, the sum of the contributions cannot be exactly 100%. The sum of the contributions ranges from 140%-67% in the six assimilation experiments. Figure 7.9 shows contribution of each parameter in all six experiments and their impacts on different variables are also shown (see legend). The positive (negative) values mean the parameters are tuned to reduce (increase) the cost function (or specific observations as legend shows).

As the dark blue bars in Figure 7.9 show, all parameters are tuned to reduce the cost function, and the minimization algorithm find a minimum (at least a local one) in the six experiments. For different constituents of the total cost function, the temperature cost reduction dominates the cost reduction, mainly caused by changes of parameter  $th2oc$ . Parameter  $acllwr$  also contributes largely to reduce the temperature cost. But its contributions are significantly reduced with decreasing nudging timescales and shows almost no contribution when the nudging timescale is smaller than 0.25 days. Parameters  $tfrc2$  and  $tdissd$ , which converge to similar values in all six experiments consistently, improve

temperature and divergence. Although not converging to similar values,  $vdiff_d$  also considerably improves the slave system in the six experiments. The other parameters show very small contributions to the cost reduction while their estimated values differ for the six experiments and even tuned into different directions. It is possible that these parameters cannot be very well observed by the observations used here and more observations or background terms are required to constrain these parameters.

#### *a. Effects of the parameters tuning long wave radiation*

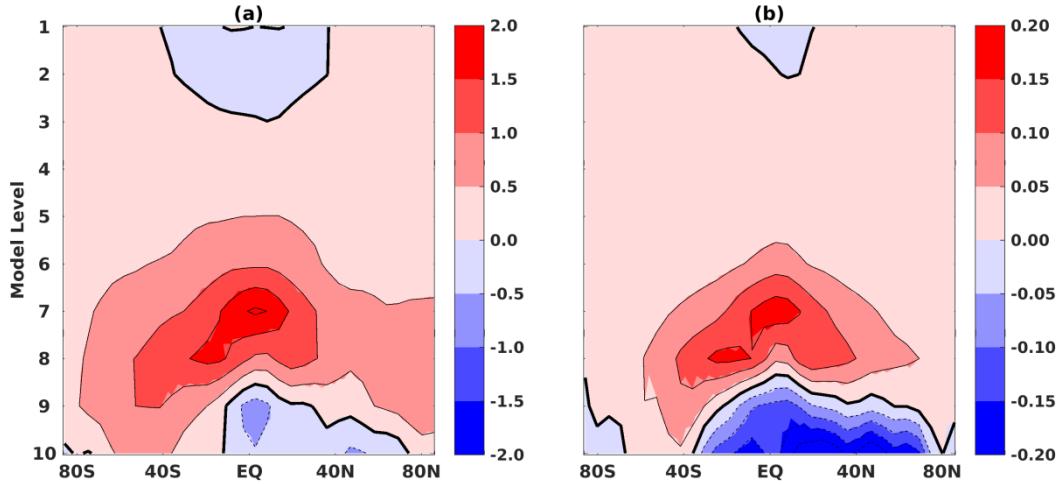


Figure 7.10. Zonally averaged temperature increments ( $^{\circ}\text{C}$ ) caused by changes of parameter  $th2oc$  in the assimilation experiments with nudging timescales of  $\tau=2.03\text{days}$  (a) and  $\tau=0.09\text{days}$  (b). The contour interval is  $0.5\text{ }^{\circ}\text{C}$  in (a) and  $0.05\text{ }^{\circ}\text{C}$  in (b).

Among all the 16 parameters, parameters  $th2oc$ ,  $acllwr$ , and  $tpofmt$  are used to tune the absorption (longwave) of water vapor and clouds. As Figure 7.9 shows, parameter  $th2oc$  contributes most to the cost function reduction, especially the temperature cost, in all the assimilation experiments. Parameter  $acllwr$  also contributes largely to reduce the temperature cost while its contributions are reduced with decreasing nudging timescales. Parameter  $tpofmt$  slightly reduces the temperature cost. In this part, the detailed effects of the three parameters are analyzed.

Figure 7.10 shows the zonally averaged temperature increments caused by changes of parameter  $th2oc$  in assimilation experiments with nudging timescales of  $\tau=2.03\text{days}$  and  $\tau=0.09\text{days}$ . Their patterns are similar, but the amplitudes are reduced to 0.1 times. Decreasing nudging timescales from  $\tau=2.03\text{days}$  to  $\tau=0.09\text{days}$  will increase the nudging strength from 0.03 to 0.4, and therefore the temperature increments are reduced in the two assimilation experiments. Change of parameter  $th2oc$  decreases the temperature in lower model levels over the land regions and the tropics while it increases the temperature in almost all the other regions. Compared to zonally averaged temperature difference as in Figure 7.5(a), the temperature increment caused by changes of parameter  $th2oc$  significantly reduces the temperature difference.

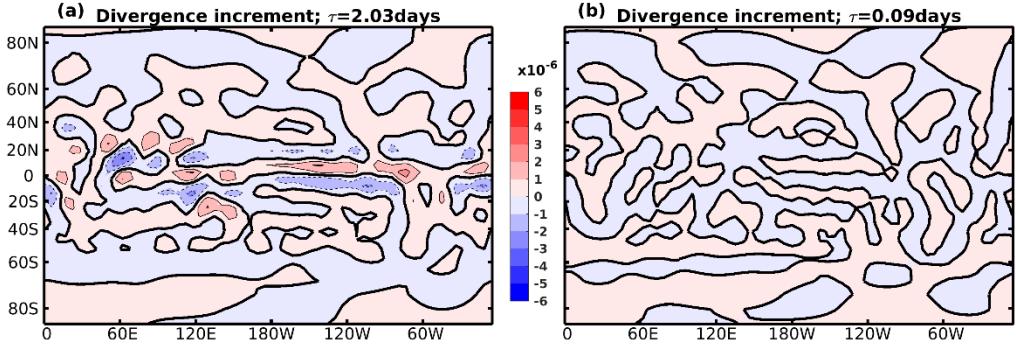


Figure 7.11. Divergence increments ( $1/s$ ) caused by changes of parameter  $th2oc$  in the assimilation experiments with nudging timescales of  $\tau=2.03\text{days}$  and  $\tau=0.09\text{days}$ .

It is worth to note that changes of the parameter  $th2oc$  also improves divergence, and the improvements are reduced with decreasing nudging timescales as shown in Figure 7.9. The divergence increments in model level 10 caused by changes of the parameter  $th2oc$  in assimilation experiments with nudging timescales of  $\tau=2.03\text{days}$  and  $\tau=0.09\text{days}$  are shown in Figure 7.11. Compared to the total divergence increments as displayed Figure 7.7(b)-(d), parameter  $th2oc$  contributes to improving divergence in the tropics in the assimilation experiment with a nudging timescale of  $\tau=2.03\text{days}$ . While in the assimilation experiments with a nudging timescale of  $\tau=0.09\text{days}$ , parameter  $th2oc$  has little effect on divergence. It is likely that parameter  $th2oc$  is informed by temperature observations and the improved temperature also improves the simulated divergence.

The zonally averaged temperature increments caused by changes of parameter  $acllwr$  are shown in Figure 7.12. In the assimilating experiments with a nudging timescale of  $\tau=2.03$  days, changes of parameter  $acllwr$  mainly reduces the warm bias in the tropics centered around level 3 as Figure 7.12(a) shows. It leads to both positive and negative effects in the other regions. The effects of parameter  $acllwr$  are significantly reduced with decreasing nudging timescales, and even the patterns of the temperature increment are changed. Parameter  $acllwr$  influences air temperature through tuning the absorption coefficient of clouds. The cloud cover in the control run and with nudging timescales of  $\tau=2.03\text{day}$  and  $\tau=0.09\text{day}$  is shown in Figure 7.13. Figure 7.13(c) shows that the small nudging timescale significantly changes the pattern of cloud cover and therefore change the sensitivity of parameter  $acllwr$ . An explanation is that: the strong nudging terms act as strong sources of temperature and moisture; they introduce more moisture into the slave system and change saturate specific humidity at the same time; the cloud cover, which is diagnosed base on threshold values of relative humidity, will also be changed. Using small nudging timescales will lead to bad synchronization considering the diagnostic variables such as cloud cover. Therefore, it is better not to use too small nudging timescales.

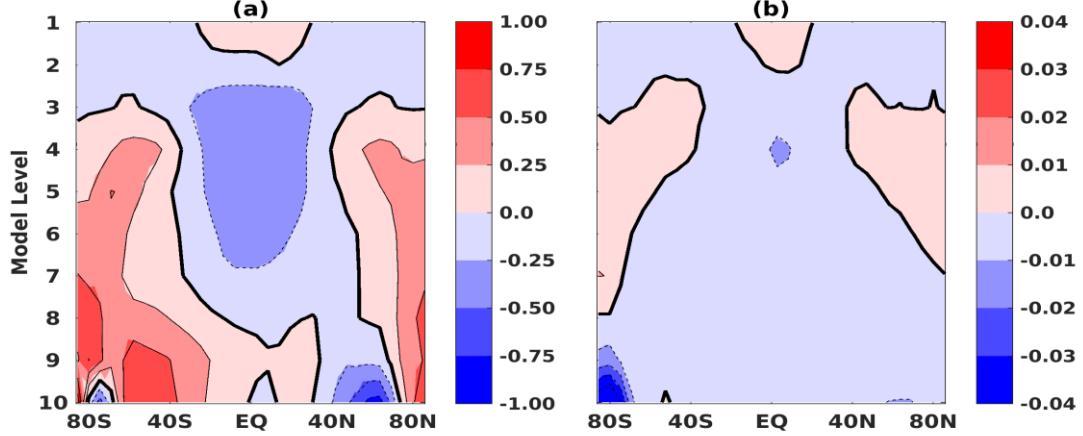


Figure 7.12. Zonally averaged temperature increments ( $^{\circ}\text{C}$ ) caused by changes of *acllwr* in the assimilating experiments with nudging timescales of  $\tau=2.03\text{days}$  and  $\tau=0.09\text{days}$ . The contour interval is  $0.25\text{ }^{\circ}\text{C}$  in (a) and  $0.01\text{ }^{\circ}\text{C}$  in (b).

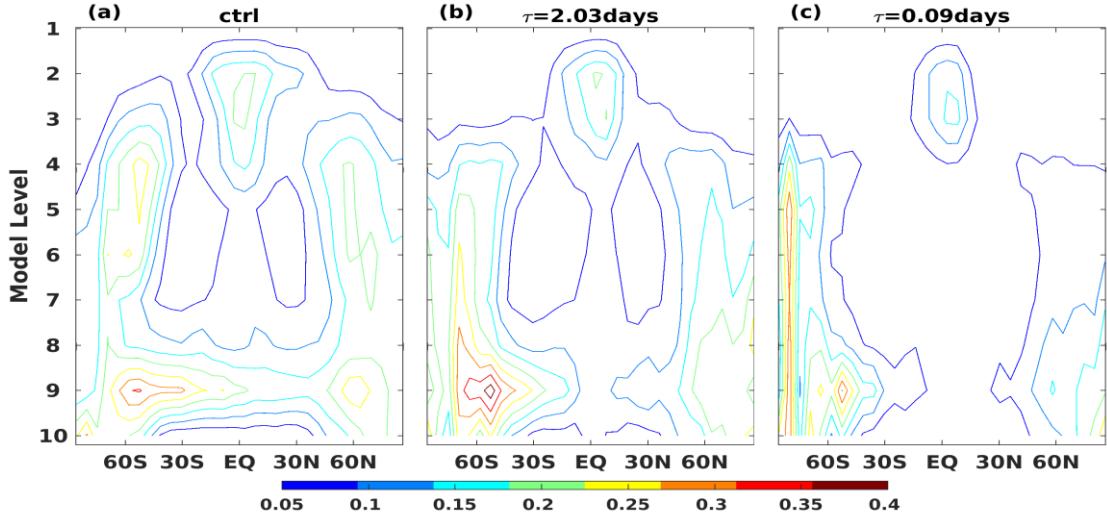


Figure 7.13. Zonally averaged cloud cover in the control run and with nudging timescales of  $\tau=2.03\text{days}$  and  $\tau=0.09\text{days}$ .

Parameter *tpofmt* tunes mean transmissivity (longwave) for a layer. Figure 7.14 shows temperature increments caused by changes of parameter *tpofmt*. As Figure 7.9 shows, parameter *tpofmt* only slightly reduces the temperature cost, and its contribution is very small compared to parameters *th2oc* and *acllwr*. The estimated values in all the assimilation experiments also vary. Parameter *tpofmt* is changed from 1. to 1.5 in the assimilation experiment with a nudging timescale of  $\tau=2.03\text{days}$  while is changed to 0 in the assimilation experiment with a nudging timescale of  $\tau=0.09\text{ days}$ . As Figure 7.14 shows, decreasing parameter *tpofmt* increases the air temperature in lower model levels while decreases air temperature in upper levels, and vice versa. But its overall effects are very small, and parameter *tpofmt* is likely not very well informed by the observations in the cost function.

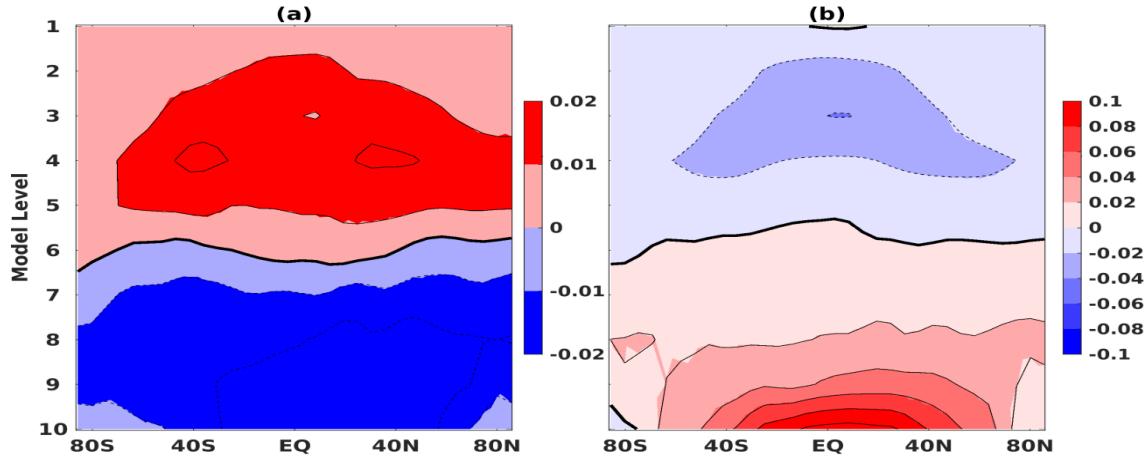


Figure 7.14. Zonally averaged temperature increments ( $^{\circ}\text{C}$ ) caused by changes of parameter  $\text{tpofmt}$  in the assimilation experiments with nudging timescales of  $\tau=2.03\text{days}$  and  $\tau=0.09\text{days}$ . The contour interval is  $0.01\text{ }^{\circ}\text{C}$  in (a) and  $0.02\text{ }^{\circ}\text{C}$  in (b).

### b. Effects of Rayleigh friction

In PlaSim, changes of Rayleigh friction directly influence divergence and vorticity in model level 01- level 02 and therefore have impacts on temperature through thermal wind balance. In all assimilating experiments, parameter  $\text{tfrc2}$  is reduced to 0.1 times of its default value as Figure 7.3 shows, which reduces the temperature cost and the divergence cost as Figure 7.9 shows. Changes of parameter  $\text{tfrc2}$  and its effects on the slave model simulation are explained below.

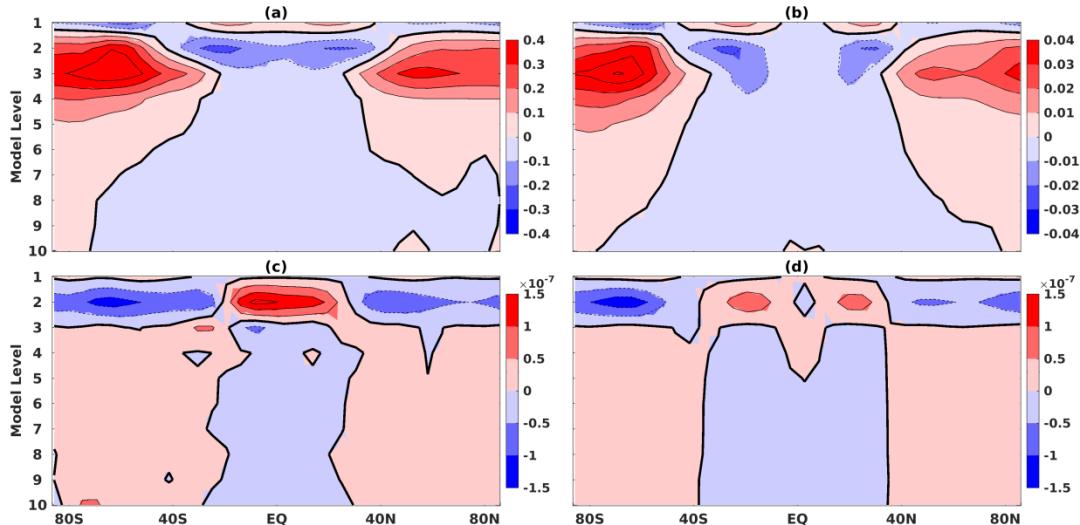


Figure 7.15. Zonally averaged temperature increments ( $^{\circ}\text{C}$ ) caused by changes of parameter  $\text{tfrc2}$  in the assimilating experiments with nudging timescales of  $\tau=2.03\text{days}$  (a) and  $\tau=0.09\text{days}$  (b). (c) and (d) are the same as (a) and (b) except that (c) and (d) are the divergence increments.

The effects of parameter  $\text{tfrc2}$  on temperature and divergence in the assimilating experiments with nudging timescales of  $\tau=2.03$  days and  $\tau=0.09$  days are shown in Figure 7.15 (a) and (b). Similar to effects of parameter  $\text{th2oc}$ , the temperature increment patterns are

similar with different nudging timescales while the amplitudes are reduced when the nudging timescale is decreased from  $\tau=2.03$  days to  $\tau=0.09$  days. As Figure 7.15(c) and (d) show, the divergence increments in both the assimilation experiments share both similar patterns and similar amplitudes, which indicates that parameter  $tfrc2$  is likely not sensitivity to the divergence observations after the synchronization.

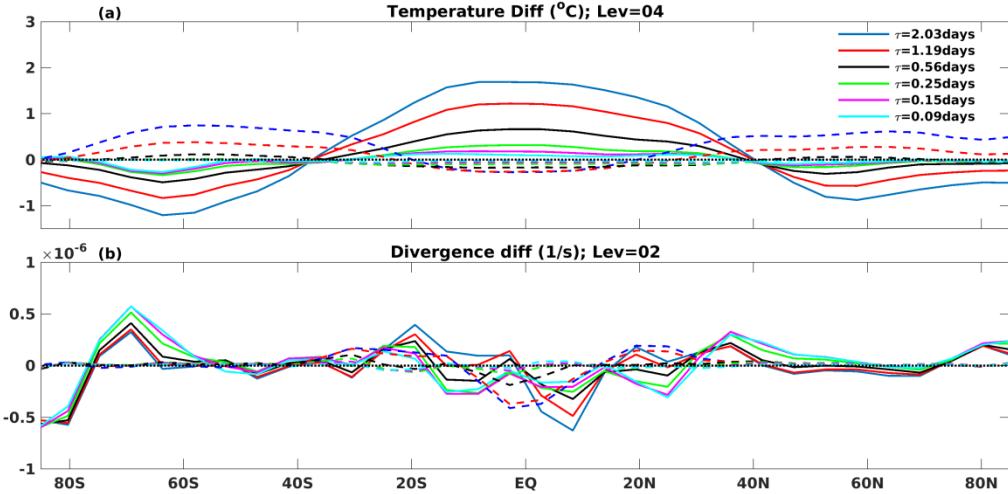


Figure 7.16. (a) Zonally averaged temperature differences (solid lines) and zonally averaged temperature increments (dashed lines) in level 04 in all the assimilating experiments. (b) Similar to (a) but for divergence and in level 02. The line colors represent the assimilation experiments with different nudging timescales as the legend indicates.

As chapter 5.2.1 shows, parameter  $tfrc2$  directly tunes the damping strength of Rayleigh friction on divergence and vorticity. Then, the changes of divergence and vorticity influence the air temperature through thermal wind balance. Changes of parameter  $tfrc2$  reduce both the temperature cost and the divergence cost. We note that changes of parameter  $tfrc2$  largely reduce the temperature cost in level 4 and the divergence cost in level 2. The temperature differences and the divergence differences between the slave models and the ERA-Interim data in the corresponding levels are shown in Figure 7.16. The increments are also shown in the dashed lines. Changes of parameter  $tfrc2$  reduce the temperature differences in all the assimilation experiments, especially in the polar region. In the tropics, parameter  $tfrc2$  slightly reduces the temperature bias. With deceasing nudging timescales, the temperature differences are reduced and the increments caused by changes of parameter  $tfrc2$  are also reduced. Compared to the temperature differences, the divergence differences seem too noisy and can hardly be reduced by changes of parameter  $tfrc2$  as Figure 7.16(b) shows. In the assimilation experiments, it is likely that parameter  $tfrc2$  is informed by temperature observations rather than the vorticity observations and the divergence observations. With nudging divergence and vorticity, sensitivities of the two observations with respect to parameter  $tfrc2$  are damped. The divergence observations and the vorticity observations cannot inform the control parameters. The usefulness of the divergence observations and the vorticity observations in the cost function will be further validated with another assimilation experiment, and the results will be discussed below.

### c. Effects of horizontal diffusion timescales

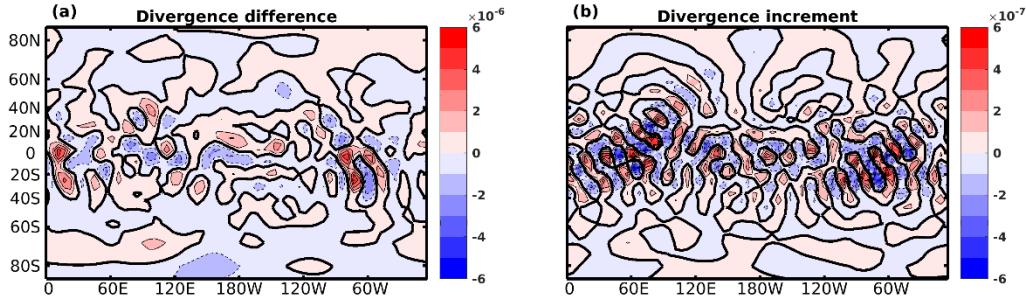


Figure 7.17. Divergence difference (a) and divergence increment (b) in level 3 caused by changes of parameter  $tdiissd$  in the assimilation experiment with a nudging timescale of  $\tau=2.03\text{days}$ .

As introduced in 5.2.2, parameters  $tdisst$ ,  $tdissz$ ,  $tdissq$  and  $tdiissd$  act to damp the prognostic variables with wavenumber larger than 15, which also make the model stable. Here, effects of parameter  $tdiissd$  in the assimilation experiment with a nudging timescale of  $\tau=2.03\text{days}$  are picked out as an example to illustrate the effects of these parameters. In all the assimilation experiments, parameter  $tdiissd$  is reduced from 0.2 days to 0.09 days and considerably reduces the divergence cost, especially in model level 3. The divergence difference and the divergence increment caused by changes of parameter  $tdiissd$  are shown in Figure 7.17. The divergence increment shows spectral ripple pattern and is one order smaller than the divergence difference. Decreasing parameter  $tdiissd$  as in all the assimilating experiments increases damping effects on divergence and makes the nonlinear system more linear. From Figure 7.17, it seems that changes of parameter  $tdiissd$  act to damp the divergence difference due to the coarse orography, such as around the Andes in southern America. To better illustrate effects of parameter  $tdiissd$ , the divergence differences and the divergence increments along 20S in level 3 are plotted in Figure 7.18. The divergence differences are large, especially in the land regions such as 60 W where the Andes located, and show wave patterns. The optimization slightly reduces these local differences by enhancing the damping effects. Parameter  $tdiissd$  are tuned to compensate these noise-like signals caused by the coarse orography of the model. The effects of parameter  $tdissz$ ,  $tdisst$ ,  $tdissq$  to their corresponding variables are similar to that of parameter  $tdiissd$ , but their sensitivities are much smaller. Several reasons may explain performance of these parameters in the assimilating experiments: (1) as Figure 7.17 displays, the divergence difference and the vorticity difference are too noise, and these parameters are tuned to damp the local noise; (2) in the slave systems, the nudging terms also act as the damping terms which largely reduce sensitivities of the horizontal diffusion parameters, and the parameters may even be tuned to negative values such as parameter  $tdisst$ . Overall, optimizing these parameters has little effect on the cost reduction.

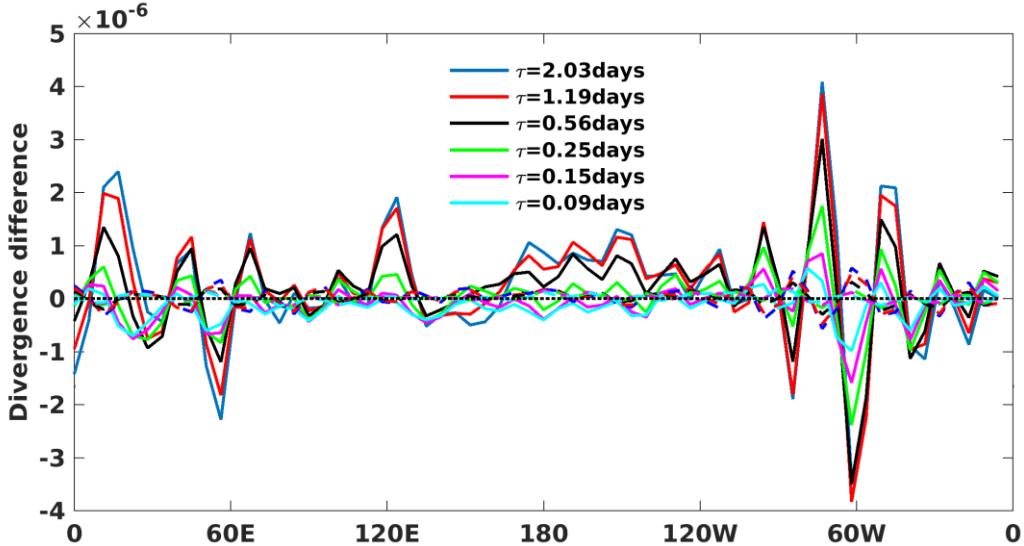


Figure 7.18. Divergence differences (solid lines) and divergence increments (dashed lines) along 20S in level 3 in all the assimilating experiments. The line colors represent the assimilating experiments with different nudging timescales as the legend indicates.

#### *d. Effects of vertical diffusion parameters*

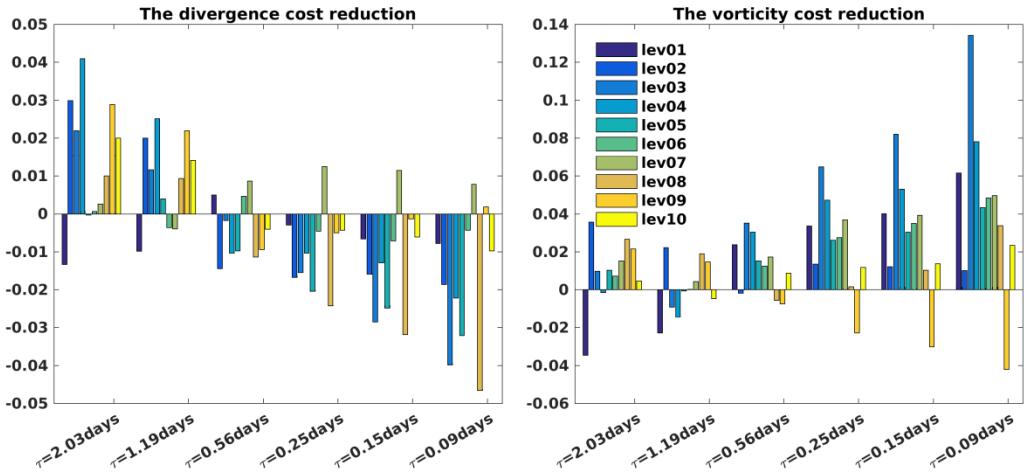


Figure 7.19. Contribution of parameter *vdiff\_lamm* to the divergence cost reduction and the vorticity cost reduction in the assimilation experiments with different nudging timescales (horizontal axis). The bar colors indicate model levels as the legend shows.

Parameters *vdiff\_b*, *vdiff\_c*, *vdiff\_d* and *vdiff\_lamm* tune the strength of vertical diffusion. In the assimilation experiments, the estimated values of these parameters vary, and their contributions to the cost function reduction are small as Figure 7.9 displays. For these four parameters, parameter *vdiff\_c* shows almost no contribution and the other three parameters mainly influence temperature, vorticity, and divergence. The performance of these parameters in the assimilation experiments is analyzed based on parameters *vdiff\_lamm* and *vdiff\_d*.

The contributions of parameter *vdiff\_lamm* to the divergence cost reduction and the vorticity cost are shown in Figure 7.19. From a default value of 160m, parameter *vdiff\_lamm* is tuned to 355.84m in the assimilating experiments with a nudging timescale of  $\tau=2.03\text{day}$

while is tuned to 107.56m in the assimilating experiments with a nudging timescale of  $\tau=0.09$ day, which leads to different contributions to the cost function reduction. In the assimilation experiments with large nudging timescales ( $\tau\geq 1.19$ days), changes of parameter *vdiff\_lamm* reduce both the divergence cost and the vorticity cost. With decreasing nudging timescales, changes of parameter *vdiff\_lamm* reduce the vorticity cost while increasing the divergence cost reduction, especially in model level 3. It seems that parameter *vdiff\_lamm* can reduce the vorticity cost more easily than the divergence cost with deceasing nudging timescales.

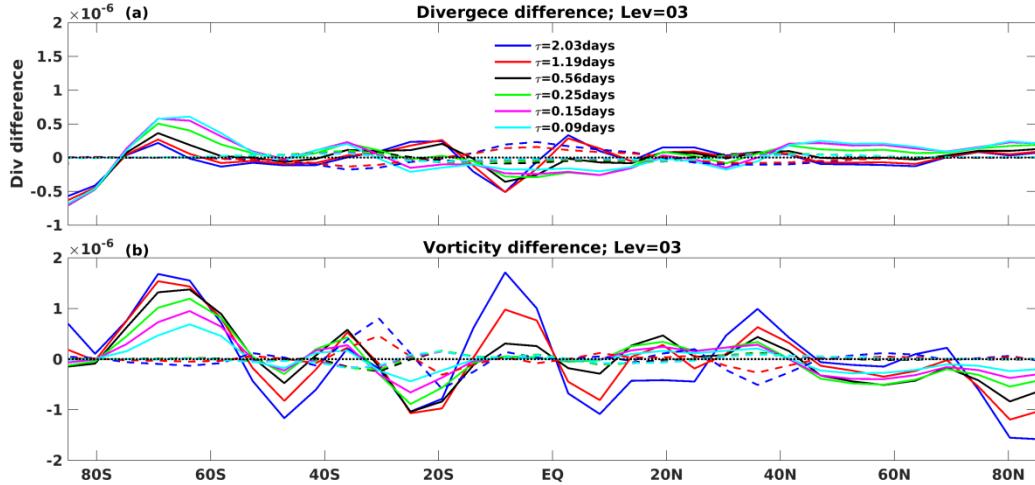


Figure 7.20. (a) Divergence differences (solid lines) and divergence increments (dashed lines) along 20S in level 3 in all the assimilating experiments. (b) Same as (a) but for vorticity. The line colors represent the assimilation experiments with different nudging timescales as the legend indicates.

The zonally averaged differences and the increments of divergence and vorticity caused by changes of parameter *vdiff\_lamm* in model level 3 are shown in Figure 7.20. Both the vorticity differences and the divergence differences show wave-like patterns along latitude. Changes of parameter *vdiff\_lamm* only slightly reduce the vorticity differences around 20S-40S and 30N-40N depending on the nudging timescales while having little effects on the divergence differences. Although changes of parameter *vdiff\_lamm* considerably reduce the vorticity cost, they only slightly reduce the vorticity differences. Their contributions to the cost function reduction are enlarged by the normalizing factors which make the vorticity cost at every level equal to 0.5. Moreover, the wave-like vorticity difference pattern along latitude indicates that the vorticity observations may not provide useful information for the optimization after synchronization. Parameter *vdiff\_lamm* is likely tuned compensate local noise signals.

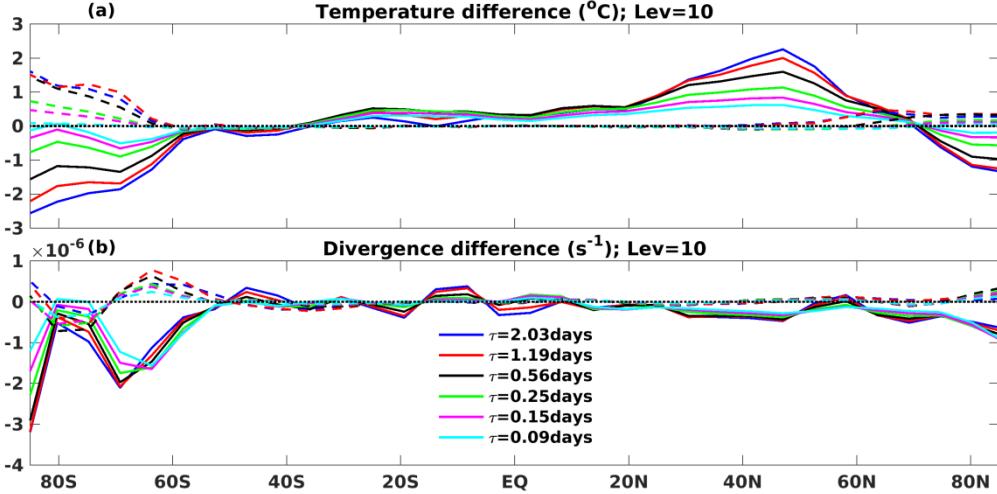


Figure 7.21. (a) Zonally averaged temperature differences (solid lines) and temperature increments (dashed lines) in level 10 caused by changes of parameter  $vdiff\_d$  in all assimilating experiments. (b) Same as (a) but for divergence. The line colors represent the assimilation experiments with different nudging timescales as the legend indicates.

In all the assimilation experiments, values of parameter  $vdiff\_d$  are significantly increased depending on the nudging timescales as Figure 7.3 shows. Changes of parameter  $vdiff\_d$  considerably reduced the divergence cost and the temperature cost in model level 10. The temperature differences and the temperature increments, the divergence differences and the divergence increments caused by changes of parameter  $vdiff\_d$  in model level 10 are shown in Figure 7.21(a) and (b) respectively. The estimated parameter  $vdiff\_d$  slightly reduces the divergence differences in 60S and also considerably reduced the temperature differences south of 60S. Parameter  $vdiff\_d$  seems to be informed by both the temperature observations and the divergence observations. Parameter  $vdiff\_d$  will be further investigated in the second group assimilating experiments.

#### e. Effects of parameters tuning shortwave radiation

Parameters  $tswr1$ ,  $tswr2$  and  $tswr3$  tune cloud optical properties. As Figure 7.3 and Figure 7.9 show, the estimated values of these parameters vary in assimilation and have little effect on the cost function reduction. The temperature increments caused by changes of these three parameters in the assimilating experiment with a nudging timescale of  $\tau=2.03\text{days}$  are shown in Figure 7.22. Compared to contributions of parameters  $th2oc$  and  $acllwr$ , which tune absorption coefficients (longwave) of water vapor and cloud, changes of these parameters have little effect on the air temperature. In PlaSim, these three parameters directly tune solar radiation at the surface and at the top of the atmosphere (TOA). They are practically tuned to get better global mean solar radiation. The divergent of the three parameters in the assimilation experiments indicate that the three parameters cannot be very well informed by the observations used in the assimilation experiments. The solar radiation observations must be included in the cost function to better inform these parameters. This Hypothesis inspires the second group of assimilating experiments, and the effects of these three parameters will be further analyzed in detail.

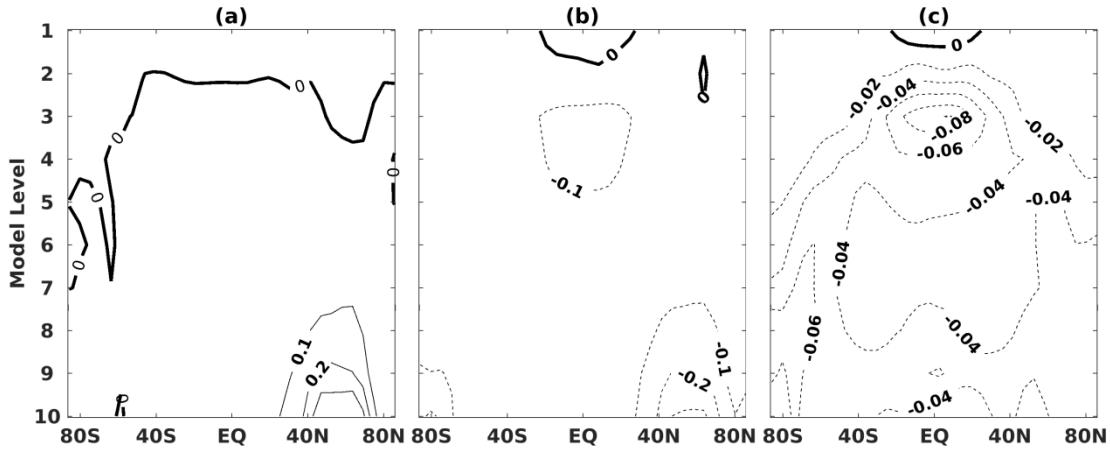


Figure 7.22. Temperature increments ( $^{\circ}\text{C}$ ) by changes of parameters  $tswr1$  (a),  $tswr2$  (b) and  $tswr3$  (c) in the assimilation experiment with a nudging timescale of  $\tau=2.03\text{days}$ .

#### 7.2.4 Assimilation of the temperature observations

As analyzed above, it seems that only temperature observations inform the control parameters and only parameters  $tfrc2$ ,  $th2oc$  and  $acllwr$  are very well observed by temperature observations. To further validate this hypothesis, another assimilation experiment is performed with only the temperature observations in the cost function and a nudging timescale of  $\tau=2.03$  days.

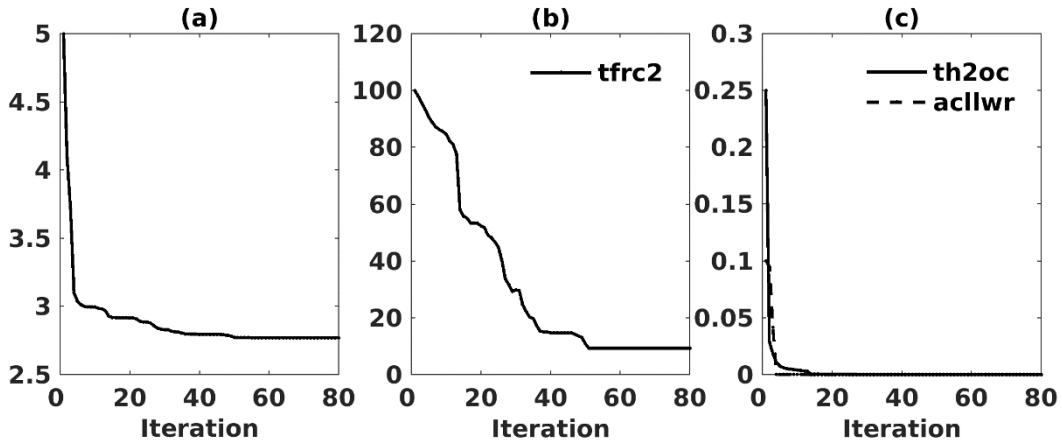


Figure 7.23. Convergence of the cost function (a), parameters  $tfrc2$  (b),  $th2oc$  and  $acllwr$  (c) in an assimilation experiment with only temperature observations in the cost function. The nudging timescale is  $\tau=2.03\text{days}$ .

Figure 7.23 displays convergence of the cost function (a), parameters  $tfrc2$  (b),  $th2oc$  and  $acllwr$  (c). The cost function is reduced by around 2.2 which is similar to the temperature cost reduction in the first group assimilation experiment with a nudging timescale of  $\tau=2.03$  days. Parameters  $tfrc2$ ,  $th2oc$  and  $acllwr$  converge to similar values as in Figure 7.3 while other parameters (not shown) are still different with that in Figure 7.3. The results validate that the three parameters  $tfrc2$ ,  $th2oc$ , and  $acllwr$  contribute most to the cost function reduction and are informed by temperature observations. While including the divergence observations and

the vorticity observations in the cost function cannot provide useful information for optimizing these parameters. The other 13 parameters cannot be very well informed by the temperature observations.

### 7.2.5 Summary

In the assimilation experiments above, the 6-hourly ERA-Interim data is used to regularize the unstable manifolds of the model simulation with nudging vorticity, divergence, temperature and specific humidity. Annually averaged vorticity, divergence, temperature and surface pressure are used in the cost function. Because the nudging terms largely reduce the cost function components and their relative importance, weighting factors are added to each constituent of the cost function, which normalizes each cost function constituent equally.

As the results show, the optimization always tunes the control parameters to reduce the cost function, especially the temperature cost and the divergence cost. The vorticity cost and the surface pressure cost is not significantly reduced. However, by carefully checking the optimization increments, we note that the divergence differences, the vorticity differences, and the surface pressure differences are too noisy to be used for parameter estimation while the temperature differences still show large-scale biases and can be used for parameter estimation. Although the divergence costs are also largely reduced in all assimilation experiments, it is mainly caused by: (1) parameters such as *th2oc* and *tfrc2* are tuned to reduce the temperature differences, and the improved temperature substantially improves divergence as Figure 7.11 shows; (2) after synchronization, the difference between the ERA-Interim data and the slave system simulation are too noise and parameter *tdissd* is tuned to damp these noisy-like signals (see Figure 7.17).

Among all the 16 parameters, only parameters *th2oc*, *acllwr* and *tfrc2* seem to be informed by the temperature observations and converge to similar values independent of the nudging timescales. Although parameter *tdissd* also converges to similar values in all the assimilation experiments, it is tuned to damp the noise-like divergence differences which are likely due to the coarse resolution of the model. Other horizontal diffusion parameters diverge and even be tuned to unphysical values. For the vertical diffusion parameters, only parameter *vdiff\_d* considerably reduced the temperature differences in the Antarctica and will be further investigated in the second group assimilation experiments. Parameters *tswr1*, *tswr2*, and *tswr3* have little contributions to the cost function reduction and are not sensitive to the observations used in the cost function. These three parameters tune cloud optical properties and will be further investigated with the shortwave radiation observations.

For the nudging timescales used for synchronization, too small nudging timescales (such as  $\tau=0.09$  days) may lead to bad synchronization regarding the diagnostic variables such as cloud cover. An explanation is that: the strong nudging terms on equations governing specific humidity and temperature will significantly increase local specific humidity and change local saturation specific humidity. The cloud cover parameterization schemes in PlaSim and the ERA-Interim reanalysis system are different. Therefore, nudging the temperature and the specific humidity too close to the ERA-Interim will not necessary simulate better cloud cover. In fact, it degrades the cloud cover field and influence estimation of parameters related to

cloud cover such as parameter *acllwr*. For the ‘maximal’ configuration, a nudging timescale of  $\tau=2.03$  days is reasonable.

### 7.3 Results for the second group assimilating experiments

#### 7.3.1 Convergence of the minimization

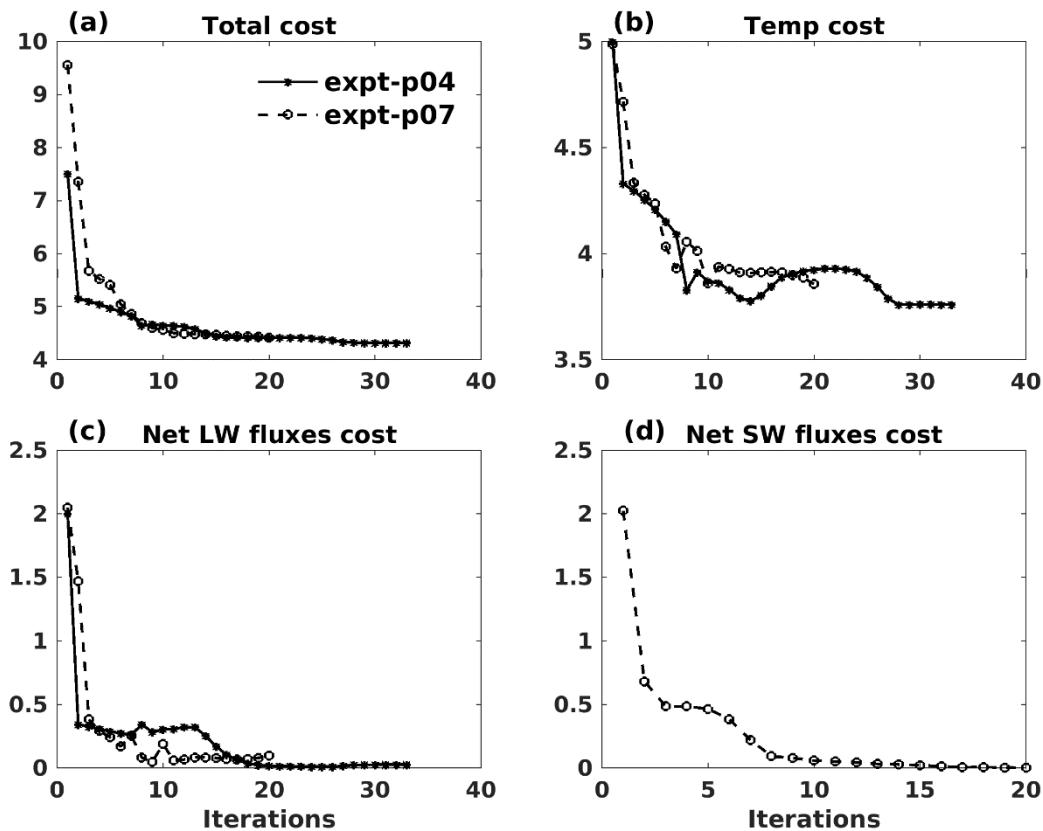


Figure 7.24. Convergence of the total cost (a), the temperature cost (b), sum of the net SW flux costs at the surface and at the TOA (c) and sum of the net LW flux costs at the surface and at the TOA vs. iteration number. The solid lines with stars are from Expt-p04 while the dashed lines with circles are from Expt-p07. In Expt-p04, net LW flux at the surface and at the TOA are not included in the cost function and therefore (d) only shows sum of the net LW flux costs at the surface and at the TOA from Expt-p07.

Figure 7.24 shows the total cost and the cost of different observations depending on iteration number. The decreases in the total cost, the temperature cost, and the flux costs are greatest during initial iterations. After the optimization, the temperature cost is reduced by 1.2 in both the assimilation experiments. In Expt-p04, the cost of global mean net LW flux at the surface and at the TOA is reduced to around 0. In Expt-p07, the costs of both the global mean net LW flux and net SW flux at the surface and at the TOA are reduced to around 0. From the perspective of the cost function, the global mean net LW flux and net SW flux biases are more easily reduced than temperature error by optimizing process parameters.

Table 7.2. The default values and the estimated values of the parameters after the

optimization.

	<b>tfrc2</b>	<b>vdiff_d</b>	<b>th2oc</b>	<b>acllwr</b>	<b>tswr1</b>	<b>tswr2</b>	<b>tswr3</b>
<b>Default</b>	100.	5.	0.25	0.1	0.0333	0.048	0.004
<b>Expt-p04</b>	8.04	62.16	0.0337	0.0499	-----	-----	-----
<b>Expt-p07</b>	8.95	40.27	0.0456	0.0349	0.0618	0.1736	0.0186

The default values and the estimated values of parameters are listed in table 7.2. After the optimization, the parameters are significantly changed. Except for parameter  $tfrc2$ , values of all the other parameters are different with that in the first group assimilation experiments. Parameter  $tfrc2$  is reduced to around 0.1 times of its default value which enhances damping effects on vorticity and divergence in upper model level. Parameters  $th2oc$  and  $acllwr$  are significantly decreased, which reduces absorption coefficient (longwave) of water vapor and cloud. But different with the first group assimilation experiments, they are larger than 0. In Expt-p07, parameters  $tswr1$ ,  $tswr2$  and  $tswr3$  are considerably increased and in their valid range as listed in table 5.1. Increasing parameter  $tswr1$  increases cloud albedo (for spectral range  $\lambda < 0.75$  um). Therefore, it allows less solar radiation reaching the surface and more reflected solar radiation to the TOA. Increasing parameter  $tswr2$  increases cloud back scattering coefficient for spectral range  $\lambda > 0.75$  um and reduces solar radiation to the surface. Increasing parameter  $tswr3$  reduces single scattering albedo for spectral range  $\lambda > 0.75$  um. Effects of these parameters on the air temperature and the radiative fluxes will be analyzed in detail below.

### 7.3.2 The air temperature improvement

#### a. The temperature improvement

As Figure 7.25 displays, the temperature increments in assimilation experiments Expt-p04 and Expt-p07 share similar patterns and amplitudes, and also similar to that in Figure 7.4(b). The temperature increments are mainly caused by changes of parameter  $th2oc$ , especially in level 6- 10. Parameters  $tfrc2$  contributes to reducing the temperature bias in model level 3 in the polar region and parameter  $vdiff_d$  reduces the temperature bias in the Antarctica as Figure 7.21 shows. The warm bias in the tropics (see Figure 7.5a) centered at level 3 is only slightly decreased by changes of parameter  $tfrc2$ . But the bias cannot be further reduced by changing parameter  $tfrc2$  because it will increase temperature errors in the polar region. In the assimilation experiment Expt-p07, the reduction of this warm bias is smaller due to changes of parameter  $tswr3$ . Increasing (decreasing) parameter  $tswr3$  will increase (decrease) the air temperature in the tropics centered at level 3 (see Figure 7.22c). In the first group assimilation experiments, parameter  $acllwr$  are also tuned to 0 to reduce the warm bias in the tropics centered at model level 3. But in Expt-p04 and Expt-p07, parameter  $acllwr$  is only reduced by 0.5 times and has little effects on temperature. This indicates that parameter  $acllwr$  is mainly informed by the flux observations. Overall, both Expt-p04 and Expt-p07 lead to similar temperature increments with that in the first group assimilation. Contributions of each parameter to the temperature cost reduction will be analyzed to explain the similarity of the temperature increments.

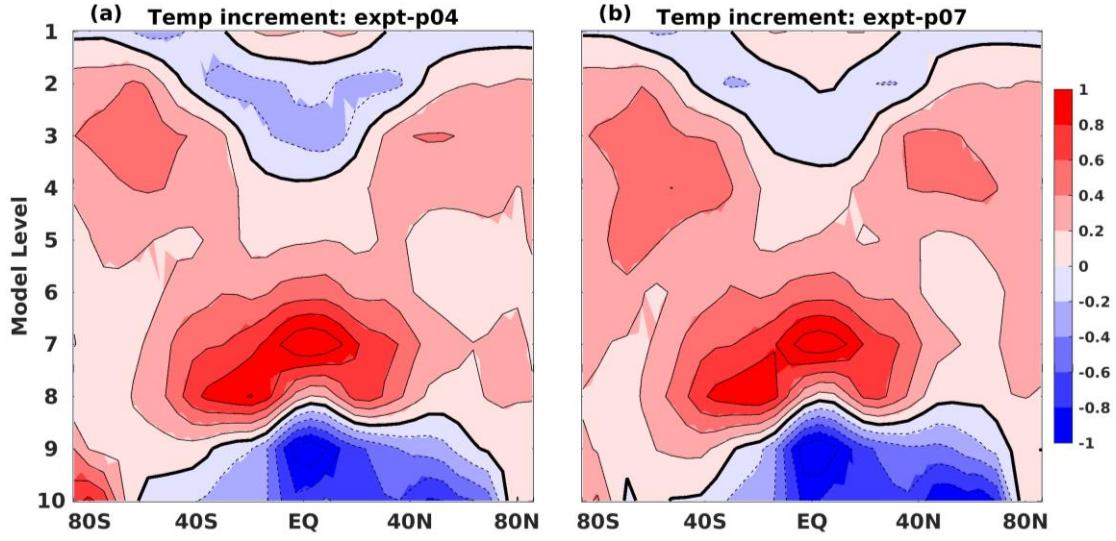


Figure 7.25. Temperature increments in both the assimilation experiments.

### b. Contributions of each parameter

Contributions of each parameter to the temperature cost reduction and the surface pressure cost reduction are approximated as in chapter 7.2.3 and are shown in Figure 7.26. The sum of the approximated contributions to the temperature cost reduction by each parameter is around 1.07 and 1.12 respectively which are similar to the temperature cost reduction in Expt-p04 and Expt-p07 as Figure 7.24(b) displays. For the temperature cost reduction, changes of parameter *th2oc* still contribute most (~70%), which lead to the similar temperature increment structures in level 6-10 in all the assimilation experiments. Both parameter *tfr2* and parameter *vdiff\_d* reduce the temperature cost by similar values in the two assimilation experiments, respectively. Parameter *aclw* has little effect on the temperature and is informed by other observations in Expt-p04 and Expt-p07. In Expt-p07, parameters *tswr1*, *tswr2* and *tswr3* play a small role in the temperature cost reduction with parameters *tswr1* and *tswr2* reduce the temperature cost while parameter *tswr3* increases the temperature cost. For the surface pressure, only parameter *tfr2* slightly reduces the cost.

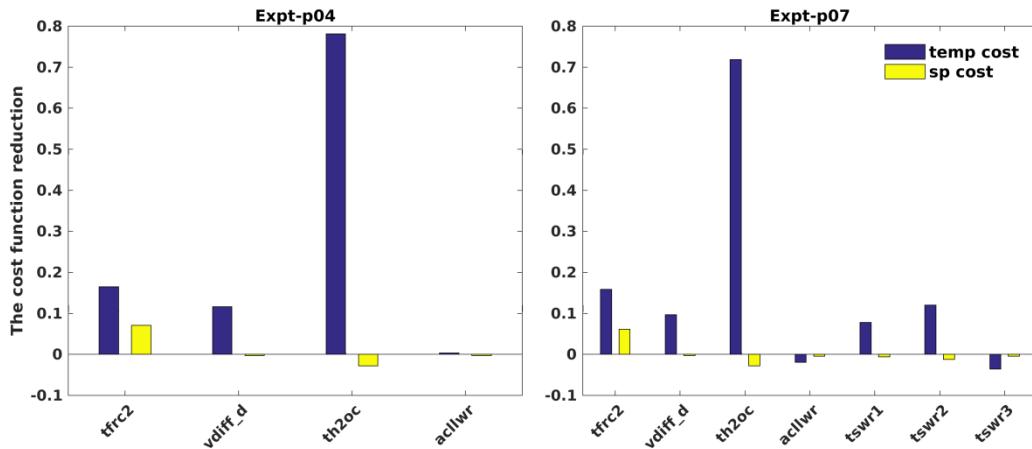


Figure 7.26. Approximated contributions of each parameter to the temperature cost reduction and the surface pressure cost reduction in the two assimilation experiments. The blue bars are

the temperature cost reduction and the yellow bars are the surface pressure cost reduction. Positive values (negative values) mean the control parameters reduced (increased) the cost by the values on y-axis.

### 7.3.3 The radiative fluxes improvement

#### a. The radiative fluxes improvement

Table 7.3. The net SW flux and net LW flux at the surface and at the TOA in the ERA-Interim data, the slave system simulation and the assimilation experiments.

	ERA-Interim	Slave system	Expt-p04	Expt-p07
Net Surface LW ( $\text{Wm}^{-2}$ )	-55.92	-26.18	-58.95	-54.38
Net Top LW ( $\text{Wm}^{-2}$ )	-245.38	-232.75	-246.90	-248.90
Net Surface SW ( $\text{Wm}^{-2}$ )	164.54	183.33	178.52	165.31
Net Top SW ( $\text{Wm}^{-2}$ )	243.75	247.90	244.11	243.59

In both Expt-p04 and Expt-p07, the biases of net flux observations are significantly reduced. The global mean net LW and SW fluxes at the surface and at the TOA in the ERA-Interim data, the slave system simulation and the two assimilation experiments are listed in table 7.3. The zonally averaged values are plotted in Figure 7.26. In the slave system, the global mean net surface LW flux, net top LW flux and net surface SW flux in the slave system have biases of  $29.7 \text{ Wm}^{-2}$ ,  $12.7 \text{ Wm}^{-2}$ , and  $19.8 \text{ Wm}^{-2}$ . For the global mean net LW flux at the surface and at the TOA, both Expt-p04 and Expt-p07 significantly reduce the biases and match very well with ERA-Interim data. For the global mean net SW flux at the surface, Expt-p04 only slightly alleviates the bias while Expt-p07 reduce the bias by  $18 \text{ Wm}^{-2}$ . In the assimilation experiments, Expt-p07 best simulates the global mean net SW flux and net LW flux at the surface and at the TOA.

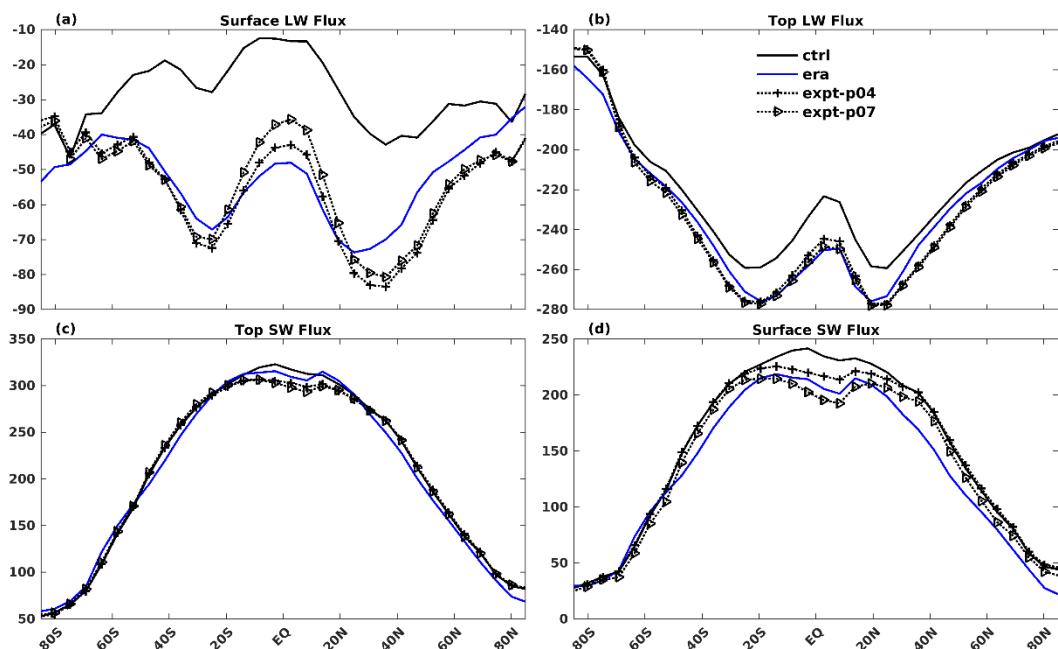


Figure 7.27. Zonally averaged net surface LW flux (a), net top LW flux (b), net top SW flux (c) and net surface SW flux (d) in the ERA-Interim data (blue lines), the slave system simulation (black lines) and the two assimilation experiments (dotted lines with markers, see legend). Negative (positive) values mean upward (downward).

As Figure 7.27(a)-(b) show, the zonally averaged net LW flux at the surface and at the TOA in the Expt-p04 and Expt-p07 are shifted downward and match the ERA-Interim data reasonably well. However, there are still errors depending on latitudes cannot be reduced. For example, Expt-p07 and Expt-p04 simulate a little more outgoing LW flux at the surface and at the TOA. But overall, the optimization significantly improves net surface LW flux and net TOA LW flux.

For net surface SW flux, the global mean value is only slightly reduced by  $4.8 \text{ Wm}^{-2}$  in Expt-p04, and the reduction is mainly in the tropics as Figure 7.27(d) shows. The slight reduction is caused by the increasing cloud cover as Figure 7.28 displays. In Expt-p04, improvement of the air temperature in the tropics substantially increases total cloud cover, which blocks solar radiation reaching the surface and therefore reduces net SW flux at the surface. While in Expt-p07, the bias of global mean net surface SW flux is reduced by  $18 \text{ Wm}^{-2}$ . The reduction is at all latitudes, especially in the tropics, as Figure 7.27(d) shows. The bias reduction of net surface SW flux is caused by: (1) the increasing cloud cover in the tropics as shown in Figure 7.28, (2) changes of parameters  $tswr1$ ,  $tswr2$ , and  $tswr3$  which tune cloud optical properties. As the results of Expt-p04 and Expt-p07 show, optimizing parameters  $tswr1$ ,  $tswr2$  and  $tswr3$  is a more efficient way to reduce the net surface SW flux bias. The net top SW flux is slightly reduced in both the assimilation experiments because the increasing cloud cover in the tropics scatters and reflects a little more solar radiation backward.

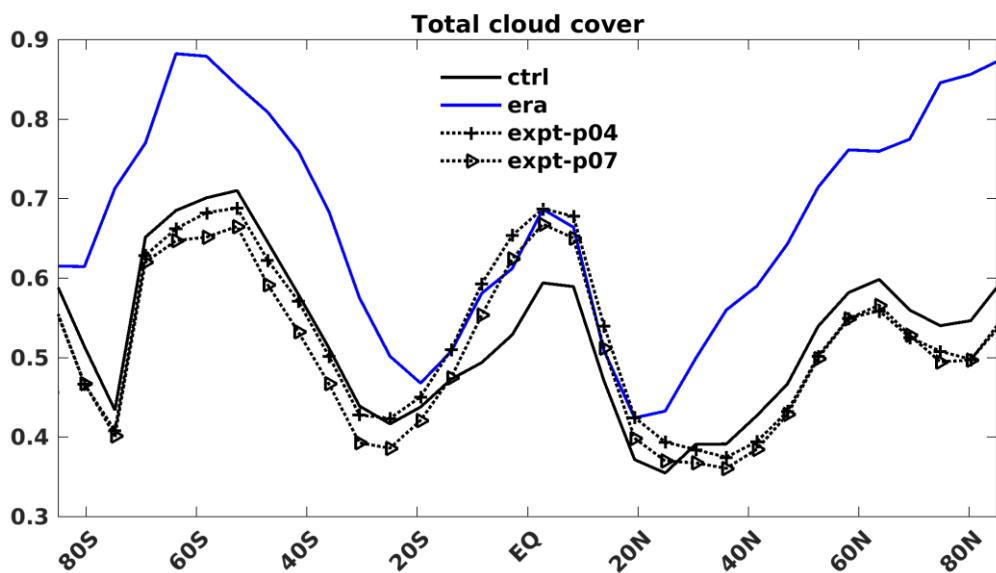


Figure 7.28. Zonally averaged totally cloud cover in the slave system simulation (black line), the ERA-Interim data (blue line) and in the two assimilation experiments (dotted lines with markers, see legend).

As the results show, the biases of net LW fluxes at the surface and at the TOA can be reasonably simulated in both Expt-p07 and Expt-p04. Due to the improvement of the air temperature in the tropics, which slightly increases the cloud cover, the bias of global mean net SW fluxes at the surface is slightly reduced in the tropics. In Expt-p07, the bias of global mean net SW fluxes at the surface is more efficiently reduced by optimizing parameters *tswr1*, *tswr2* and *tswr3*. Effects of each parameter on net LW flux and net SW flux are shown below.

### ***b. Effects of each parameter***

To better understand how the parameters contribute to improving the global mean net LW flux and net SW flux, approximated contributions of each parameter to the global mean net SW and net LW fluxes changes are shown in Figure 7.29.

In both the assimilation experiments, parameters *tfrc2* and *vdiff\_d* almost have no effects on net SW flux and net LW flux at the surface and at the TOA, and they are informed by the air temperature observations in the optimization. The rest parameters contribute significantly to the flux cost reduction. Parameters *th2oc* and *acllwr* contribute to net LW flux improvement and parameters *tswr1*, *tswr2* and *tswr3* contribute net SW flux improvement. The sum of the contributions shown in Figure 7.29 is close to the flux bias reduction listed in Table 7.3, which indicates the linear approximation can reasonably represent contributions of each parameter.

As Figure 7.28(a) displays, parameter *th2oc* alone significantly increases the global mean net surface LW flux by  $-30 \text{ Wm}^{-2}$  (negative indicates outgoing). As for net top LW flux, parameters *th2oc* and *acllwr* contribute to reducing the bias by  $-10 \text{ Wm}^{-2}$  and  $-4 \text{ Wm}^{-2}$ , respectively. And their contributions are slightly different in Expt-p07 and Expt-p04. The values of parameters *th2oc* and *acllwr* are decreased in the optimization, which reduces absorption coefficients (longwave) of water vapor and cloud and therefore increases outgoing LW flux at the surface and at the TOA. By optimizing parameters *th2oc* and *acllwr*, the global mean net LW flux at the surface and at the TOA can be well very well simulated. Tuning parameters *th2oc* is more efficient considering the air temperature observations and the LW flux observations.

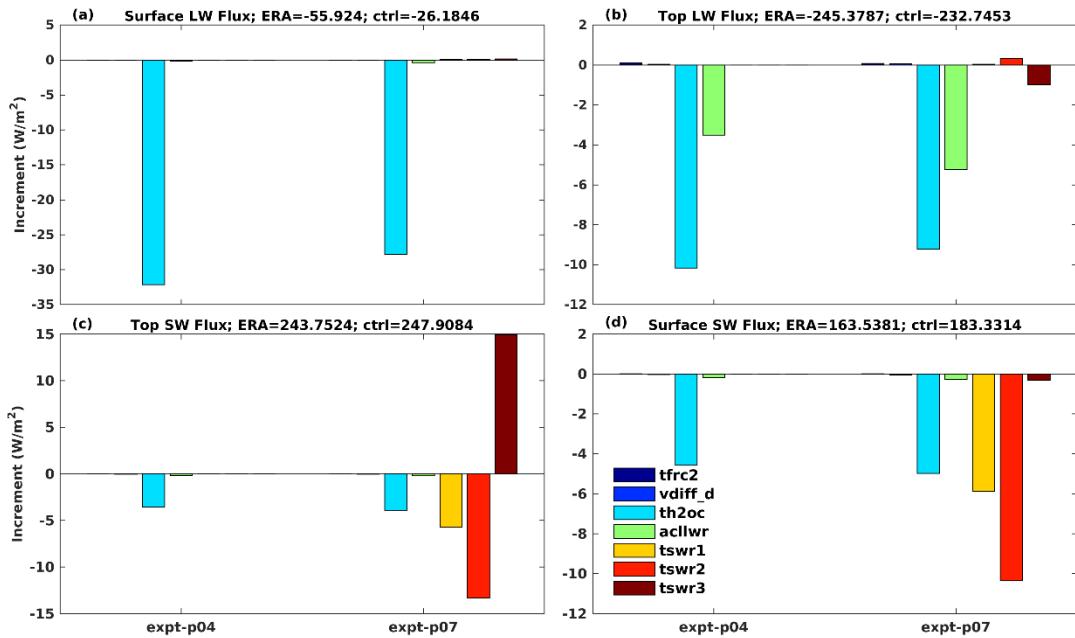


Figure 7.29. Approximated contributions of each parameter (see legend in d) to net surface LW flux (a), net top LW flux (b), net top SW flux (c) and net surface SW flux (d). Positive values (negative values) mean the control parameter increases (decreases) the corresponding variables. The corresponding values of ERA-Interim data and the slave system simulated values are on top of each subplot.

For net SW flux as Figure 7.28(c)-(d) show, parameter *th2oc* slightly reduces net surface SW and net top SW flux due to the increasing cloud cover in the tropics. Parameters *tswr1*, *tswr2*, and *tswr3* dominate the changes of net SW flux at the surface and at the TOA. For net surface SW flux which has a global mean bias of  $19.8 \text{ Wm}^{-2}$ , parameters *tswr1* and *tswr2* are tuned to reduce the bias by  $-6 \text{ Wm}^{-2}$  and  $-10 \text{ Wm}^{-2}$  while parameter *tswr3* has little effect. But changes of parameters *tswr1* and *tswr2* also significantly decrease net top SW flux by  $-5 \text{ Wm}^{-2}$  and  $-13 \text{ Wm}^{-2}$ , and degrade net top SW flux. Parameter *tswr3* is tuned to increase net top SW flux by  $15 \text{ Wm}^{-2}$ , which compensates the negative effects of parameters *tswr1* and *tswr2*.

Figure 7.30 displays the spatial distributions of the net surface (a-c) SW flux and net top (d-f) SW flux caused by changes of parameters *tswr1*, *tswr2* and *tswr3*. Parameters *tswr1* (a) and *tswr2* (b) reduce net surface SW flux globally by increasing albedo (for spectral range  $\lambda < 0.75 \text{ um}$ ) and backscattering coefficient (for spectral range  $\lambda > 0.75 \text{ um}$ ) of clouds. The change of parameter *tswr3* also decreases surface SW flux in most regions of the tropics while increasing it between 20S-60S and 20N-60N over the ocean region. But, effects of parameter *tswr3* on the global mean surface SW flux is very small as Figure 7.29(d) shows.

The global mean net top SW flux is reasonably well in the slave system simulation. However, changes of parameters *tswr1* (d) and *tswr2* (e) largely reduce net top SW radiation globally by increasing the reflected and backscattered SW flux to the TOA. Therefore, changes of parameters *tswr1* and *tswr2* degrade the global mean net top SW flux. The negative effects of parameters *tswr1* and *tswr2* on net top SW flux are compensated by changes of parameter *tswr3* which reduce the single scattering albedo. Overall, the

optimization efficiently improves the global mean net SW flux at the surface and at the TOA by tuning parameters *tswr1*, *tswr2*, and *tswr3*.

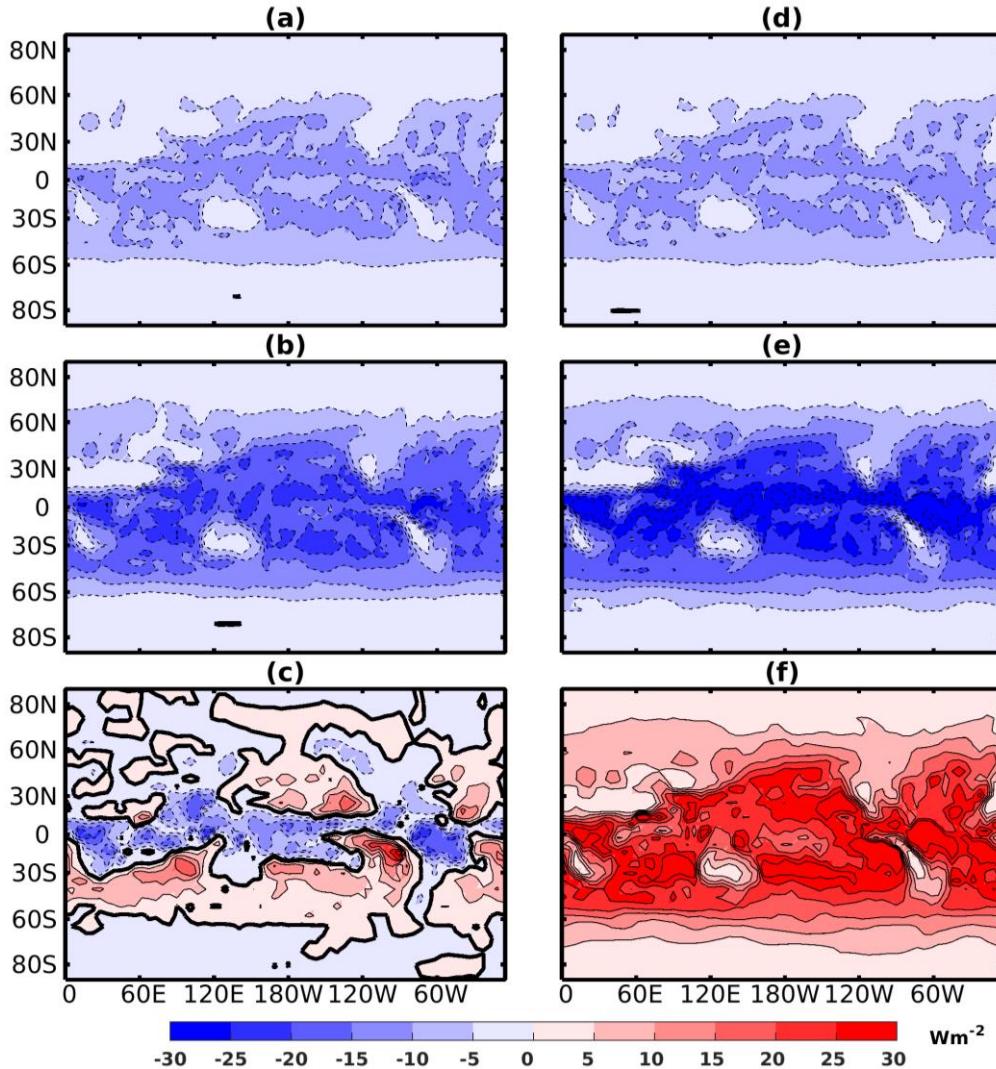


Figure 7.30. Increments of net surface SW flux (a)-(c) and net top SW flux (d)-(f) caused by changes of parameters *tswr1* (a,d), *tswr2* (b,e) and *tswr3* (c,f).

Based on the analysis above, the improvements of the model simulation caused by the optimization can be briefly explained as follow. For the net LW flux at the surface and at the TOA, the slave system simulates less outgoing LW flux both at the surface and at the TOA than the ERA-Interim data. The optimization increases net outgoing LW flux at the surface by decreasing absorption coefficient of water vapor (*th2oc*), which efficiently reduces back LW radiation to the surface. For net LW flux at the TOA, it is increased by decreasing absorption coefficients of water vapor (*th2oc*) and clouds (*acllwr*), which allows more outgoing longwave radiation. As for net SW flux at the surface, the slave system simulates more net SW flux at the surface than the ERA-Interim data. The bias is reduced by increasing albedo (*tswr1*, for spectral range  $\lambda < 0.75 \text{ um}$ ) and backscattering coefficient (*tswr2*, for spectral range  $\lambda > 0.75 \text{ um}$ ) of clouds. At the same time, the changes of albedo (for spectral range  $\lambda < 0.75 \text{ um}$ ) and backscattering coefficient (for spectral range  $\lambda > 0.75 \text{ um}$ ) of clouds increase reflected SW

flux to the TOA and makes net SW flux at the TOA worse. The negative effects on net SW flux at the TOA is compensated by reducing single scattering albedo ( $tswr3$ , for spectral range  $\lambda > 0.75 \text{ um}$ ) of clouds.

Overall, by optimizing parameter  $th2oc$  and  $acllwr$  which tunes absorption coefficients of water vapor and clouds, the global mean net LW flux at the surface and at the TOA can be very well simulated. The global mean net SW flux at the surface and at the TOA can also be very well simulated by optimizing parameters  $tswr1$ ,  $tswr2$  and  $tswr3$  which tune cloud optical properties. As for the air temperature, its improvement is mainly caused by changes of parameter  $th2oc$ . Parameter  $tfrc2$  improves temperature in upper model levels through thermal wind relation and parameter  $vdiff_d$  also slightly improves temperature in the Antarctica.

## 7.4 Evaluation with the free model

In the assimilation experiments with the slave systems, the control parameters are tuned to reduce the cost function. However, the nudging terms still retain in the slave system after the optimization and act as external forcing to push the slave system close to the ERA-Interim data. As Figure 7.1 shows, the nudging terms are large compared with the tendency terms. The purpose of the assimilation experiments is to find optimal values of the control parameters which reduce bias of the free model (without the nudging terms). Therefore, the effects of the optimized parameters on the free model must be evaluated before we draw the conclusion that the method can be applied to calibrating the climate model. Three model simulations are performed with the estimated parameters from the assimilation experiments Expt-p04, Expt-p07, and expt-p16 (the first group assimilation experiments with a nudging timescale of  $\tau = 2.03\text{days}$ ). We use Expt-p04f, Expt-p07f, and expt-p16f to represent the three free model simulations respectively and the model simulation with the default values of parameters is called control run (ctrl in the figures). The model is integrated for 3 years with the last 2-year data for comparison. Although some parameters such as  $tdisst$  is tuned to negative values in the assimilation experiment expt-p16, the free model simulation expt-p16f with the estimated parameters is still stable. First, the net SW flux and net LW flux at the surface and at the TOA which are significantly improved in the assimilation experiments Expt-p04 and Expt-p07 will be evaluated. Second, the air temperature which also plays an important role in informing the parameters will be analyzed. At last, the effects on other variables which are not explicitly used in the cost function such as zonal wind, cloud cover and specific humidity will be shown.

### 7.4.1 The Radiative fluxes

The global mean radiative fluxes in the ERA-Interim data and the four free model simulations are listed in table 7.4. The meridional distributions of the annual mean difference between the four model simulations and the ERA-Interim data are shown in Figure 7.31. In both Expt-p04f and Expt-p07f, net surface LW flux and net top LW flux are significantly improved while being seriously degraded in expt-p16f. Figure 7.31 (a) and (b) show that the control run simulates less outgoing surface LW flux and outgoing top LW flux than the ERA-Interim data at all latitudes. Both Expt-p04f and Expt-p07f improve net surface LW and

net top LW flux almost at all latitudes. However, expt-p16f seems to over fit the ERA-Interim data and simulates much more outgoing LW flux at the surface and at the TOA. In the assimilation experiments expt-p16 with a nudging timescale  $\tau=2.03$  days, parameters *acllwr* and *th2oc* are tuned to around 0 to reduce the temperature cost regardless of LW flux at the surface and at the TOA. The estimated parameters *acllwr* and *th2oc* make clouds and water vapor almost transparent to longwave radiation and therefore seriously degrades LW fluxes both at the surface and at the TOA. While in assimilation experiments Expt-p04 and Expt-p07, parameter *th2oc* is observed by temperature, net surface LW flux, and net top LW flux while parameter *acllwr* is mainly observed by net top LW flux. They are tuned to finite values to reduce the total cost. Therefore, it is critical important to include net LW flux at the surface and at the TOA in the cost function to optimize parameters which tune absorption coefficients of clouds and water vapor.

Table 7.4. Radiative fluxes, surface latent heat flux and surface sensible heat flux in the ERA-Interim data and the model simulations with parameters from different assimilation experiments. Negative values represent outgoing.

	Net Top LW (Wm <sup>-2</sup> )	Net Surface LW (Wm <sup>-2</sup> )	Net Top SW (Wm <sup>-2</sup> )	Net SW (Wm <sup>-2</sup> )	Surface flux (Wm <sup>-2</sup> )	heat	Surface heat flux (Wm <sup>-2</sup> )
ERA-Interim	-245.38	-55.92	243.75	164.54	-81.58		-17.46
Ctrl	-228.74	-31.79	239.33	176.01	-103.52		-26.09
Expt-p04f	-240.81	-63.19	239.58	176.14	-87.42		-24.10
Expt-p07f	-243.13	-59.62	239.21	161.00	-80.29		-22.47
Expt-p16f	-259.71	-110.73	244.03	190.46	-72.87		-19.47

For net SW flux at the TOA, both Expt-p04f and Expt-p07f don't change the global mean values while expt-p16f increases net SW flux by 4.7 Wm<sup>-2</sup>. Compared to net LW flux improvement as Figure 7.31(a) and (b) show, changes of net SW flux at the TOA is not significant. As Figure 7.31(d) shows, expt-p16f increases net SW flux at low latitudes and middle latitudes, and it degrades net SW flux at middle latitudes. Therefore, expt-p16f doesn't robustness improve net SW flux at TOA. From the global mean perspective, the model-simulated net SW flux at the TOA is reasonable well. However, there are still errors depending on latitudes which are seldom reduced with the optimized parameters.

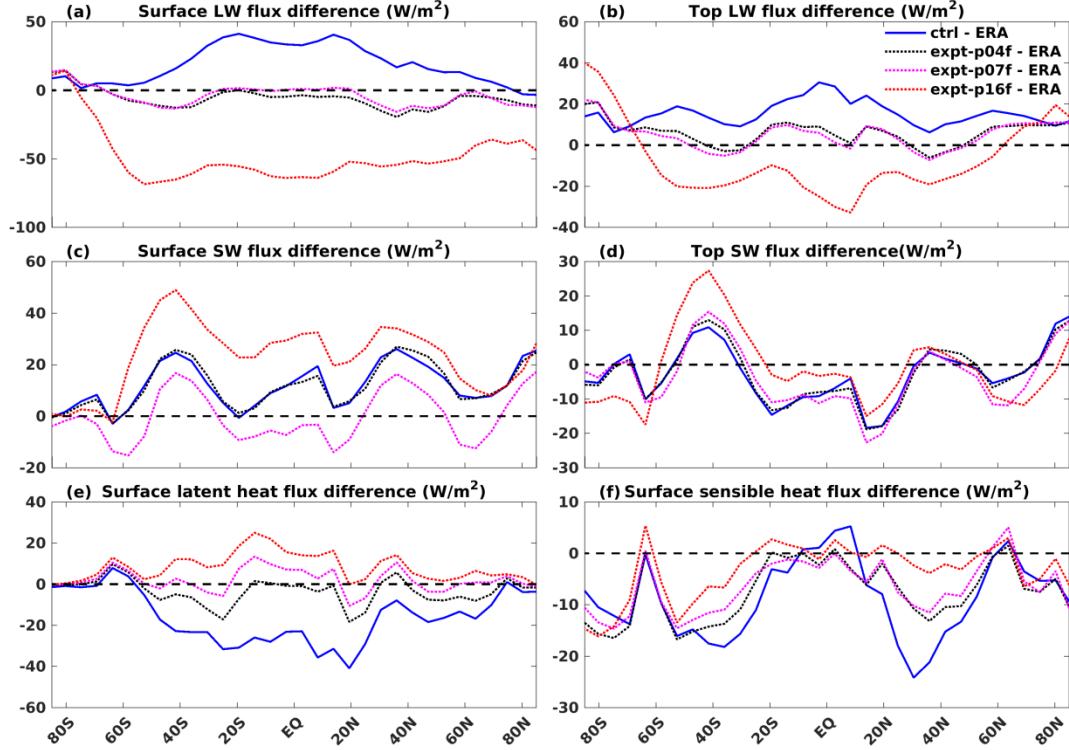


Figure 7.31. Meridional distributions of the annual mean difference of surface LW flux (a), top LW flux (b), surface SW flux (c), top SW flux (d), surface latent heat flux (e) and surface sensible heat flux between the model simulations and the ERA-Interim data. The line style and line color represent the model simulations with optimized parameters from different assimilation experiments (see legend).

For net surface SW flux as Figure 7.31(c) shows, the positive bias of global mean value is reduced only in Expt-p07f while there are still errors depending on latitudes. Expt-p04f doesn't change the bias of net surface SW flux as in the assimilation experiment Expt-p04. In the assimilation Expt-p04, the reduction of net surface SW flux is achieved by the increasing the total cloud cover in the tropics while total cloud cover changes are very small in Expt-p04f as Figure 7.36(c) shows. One of the important reasons is that the specific humidity cannot be improved in the model. Although surface latent heat flux and surface sensible heat flux are not included in the cost function, they are also considerably improved in Expt-p07f and Expt-p04f due to the improvement of temperature and radiative fluxes.

Overall, with the optimized parameters from assimilation experiments, the free model simulations Expt-p04f and Expt-p07f improve the global mean radiative fluxes in a way similar to that in the corresponding assimilation experiments: Expt-p04f significantly improves net LW flux at the surface and at the TOA while has little effect on net SW flux at the surface and at the TOA; Expt-p07f improves both net LW flux and net SW flux significantly. Surface latent heat flux is also improved, and surface sensible heat flux is slightly improved. However, in the free model simulation expt-p16f, radiative fluxes are much worse than the control simulation. This is because that: (1) parameters *th2oc* and *acllwr* are tuned to 0 to reduce the temperature cost without regarding net SW flux and net LW flux, which makes clouds and water vapor transparent to longwave radiation and significantly

degrades net LW flux at the top and at the surface; (2) parameters  $tswr1$ ,  $tswr2$ , and  $tswr3$  are not very well informed by the observations in the first group assimilation experiments, changes of these parameters significantly degrade net SW flux at the surface. Therefore, it is important to include the LW and SW fluxes in the cost function when optimizing the parameters related to clouds and water vapor properties.

#### 7.4.2 Temperature and zonal wind

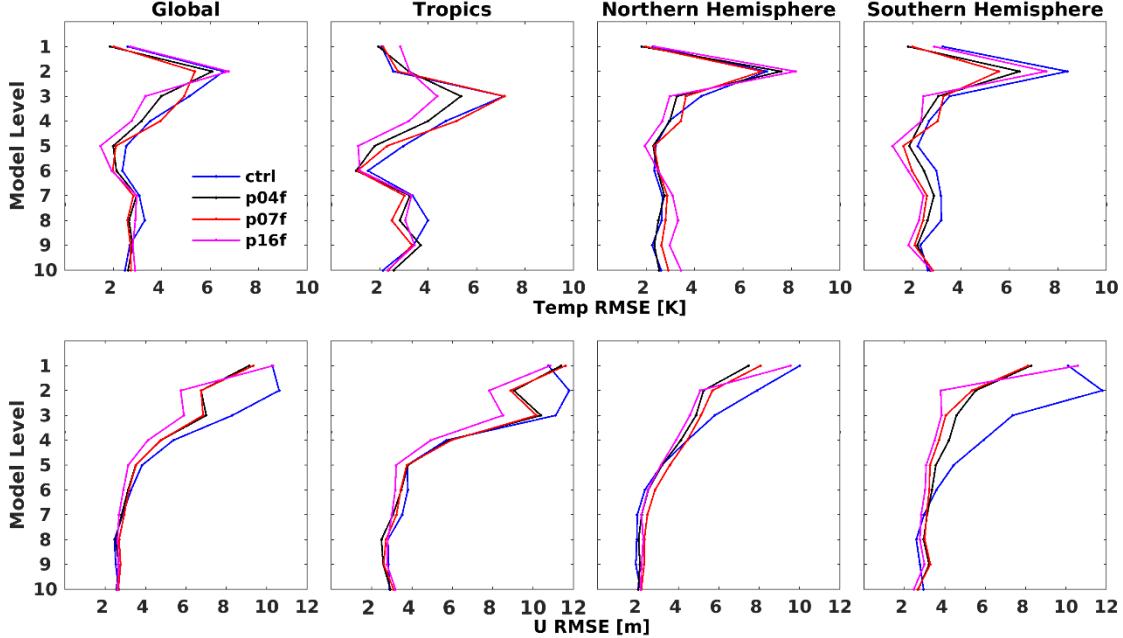


Figure 7.32. The root mean square error of temperature (upper panels) and Zonal wind (lower panels) for different geographical regions. The RMSE is computed with the 4 model simulations (see legend) and the ERA-Interim data.

The air temperature is important observations in all the assimilation experiments which inform parameters such as  $tfrc2$ ,  $vdiff_d$ , and  $th2oc$ . Although not explicitly included in the cost function, the Zonal wind can be changed by optimized parameter  $tfrc2$  and therefore is evaluated here. Figure 7.32 shows the root mean square error of temperature and Zonal wind for a global, the tropics, the northern hemisphere and the southern hemisphere average in the four free model simulations.

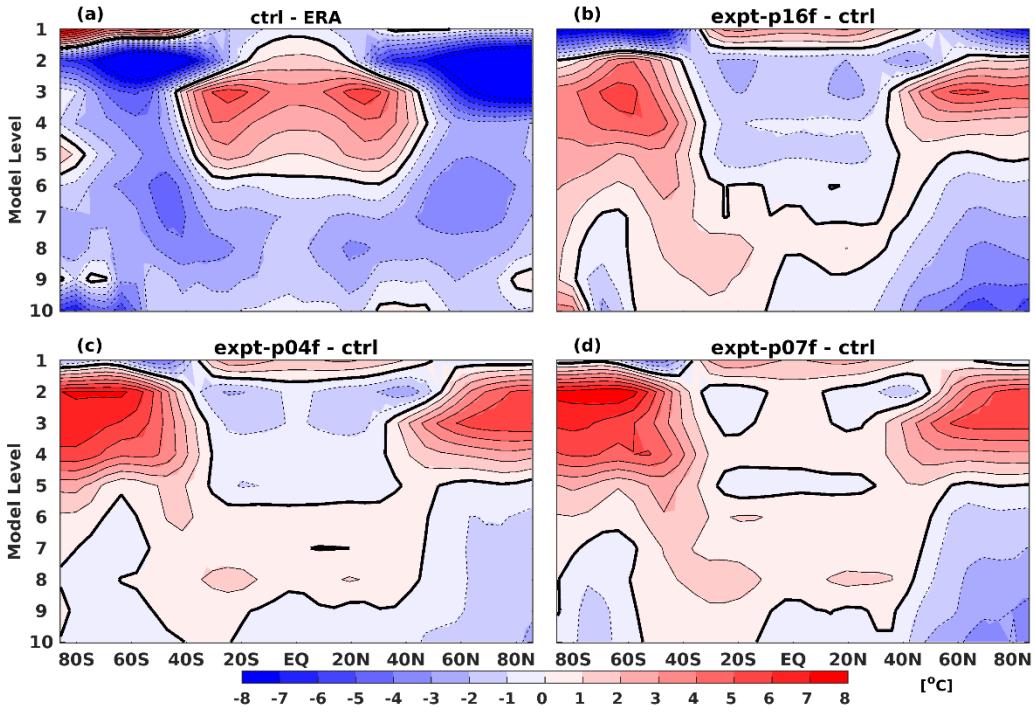


Figure 7.33. (a) Annual mean of zonally averaged temperature difference between the control run and the ERA-Interim data. Annual mean of zonally averaged temperature differences between the model simulations with optimized parameters from different assimilation experiments and the control run (b-d).

From a global perspective, the three simulations with optimized parameters reduce RMSE for both temperature and Zonal wind. The estimated parameters significantly reduce the Zonal wind RMSE in model level 2-4 where the control run simulates too strong westerlies. In lower model levels, the estimated parameters slightly increase the zonal wind RMSE in both the northern and the southern hemisphere while decreasing it in the tropics. Globally, the temperature shows an error reduction for all three simulations. However, the error reduction varies in different geographical regions and different simulations. All three experiments reduce the temperature RMSE in the tropics and in the southern hemisphere while having both positive and negative effects in the northern hemisphere. As Figure 7.33 shows, the temperature RMSE reduction in model level 2-5 is mainly caused by: (a) temperature increases in the polar region, which exist in all 3 simulations with optimized parameters, due to changes of parameter  $tfr2$ ; (2) temperature decreases in the tropics in Expt-p04f and expt-16f due to changes of parameter  $acllwr$ . In Expt-p07f, temperature decreases in the tropics centered at level 3 due to changes of parameter  $acllwr$  are compensated by changes of parameter  $tswr3$ . Parameter  $th2oc$ , which contributes most to the temperature cost reduction, contributes to reducing the cold bias in the tropics in all three simulations.

The zonal wind is mainly influenced by changes of parameter  $tfr2$ . In the assimilation experiments, parameter  $tfr2$  is informed by temperature through the thermal wind relation. The Rayleigh friction timescale  $tfr2$  is reduced from 100 days to around 10 days in the assimilation experiments as listed table 7.1, which lead to similar zonal wind patterns in the three free model simulations with optimized parameters. Therefore, zonal wind from Expt-p07f is shown for comparison. Zonally averaged zonal wind and standard deviation for

the ERA-Interim data, the control run and Expt-p07f are shown in Figure 7.34. The control run simulates too strong westerlies in both hemispheres. The optimized parameter  $tfr2$  strongly enhances the damping effect on divergence and vorticity. Therefore, the strength of westerlies is reduced and at the same time, the variance is reduced. However, it seems that the locations of the maximum zonal wind in the southern hemisphere and the northern hemisphere are not changed.

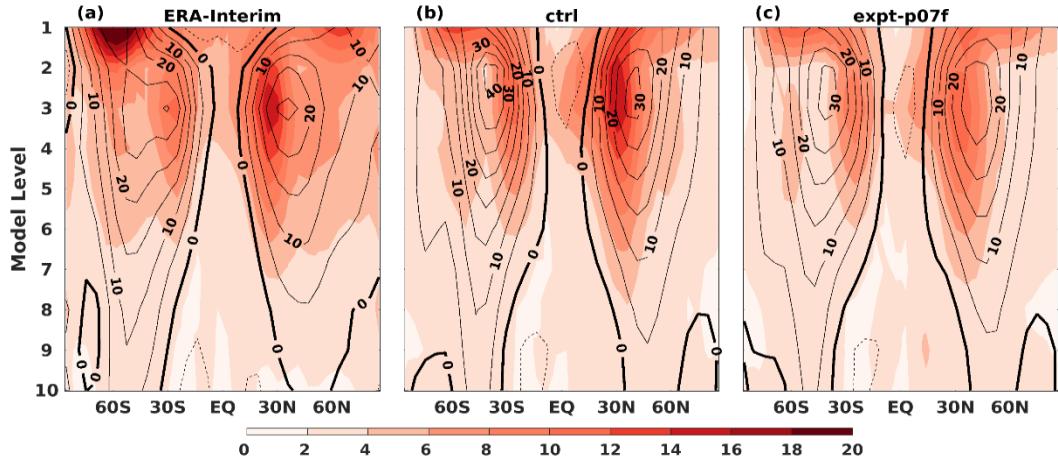


Figure 7.34. (a) Annual mean of zonal wind for the ERA-Interim data (a), the control simulation (b) and Expt-p07f (c).

#### 7.4.3 Specific humidity, precipitation and cloud cover

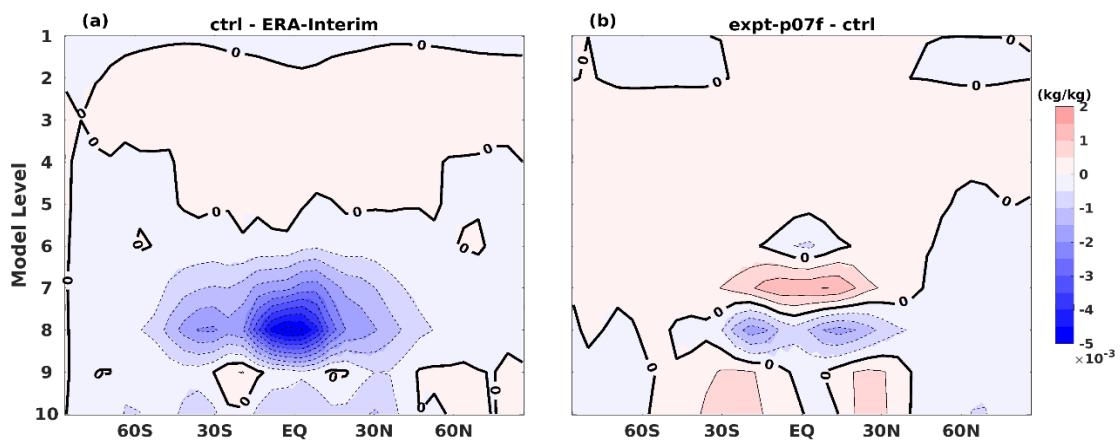


Figure 7.35. (a) Specific humidity difference between the control simulation and the ERA-Interim data, (b) specific humidity difference between Expt-p07f and the control simulation. The contour interval for the dashed and solid lines is  $0.5 \times 10^{-3}$  kg/kg.

In the control simulation, the PlaSim simulates less moisture than the ERA-Interim data as Figure 7.35(a) displays. One of the possible reasons is that the simulated temperature is too low which allows less moisture in the atmosphere. The optimized parameters increase the air temperature, especially in the tropics. But globally, the specific humidity in Expt-p07f is reduced by around 1.8%. As Figure 7.35(b) shows, the specific humidity is reduced at the tropics in model level 8 and at the northern hemisphere, which is likely because of model

deficiency. The moisture reduction in the northern hemisphere may be responsible for the temperature decreases in the northern hemisphere as Figure 7.34(b)-(d) show. The reduction of the specific humidity and the decreasing of parameter  $th2oc$  reduce the absorptivity of longwave radiation by water vapor. As Figure 7.36(c) shows, the total cloud cover is also reduced in the northern hemisphere and therefore decreases absorptivity of longwave radiation by clouds. In fact, the absorptivity of longwave radiation by clouds is also reduced by decreasing parameter  $acllwr$ . All these factors lead to the reducing temperature in the northern hemisphere and degrade the air temperature in the northern hemisphere.

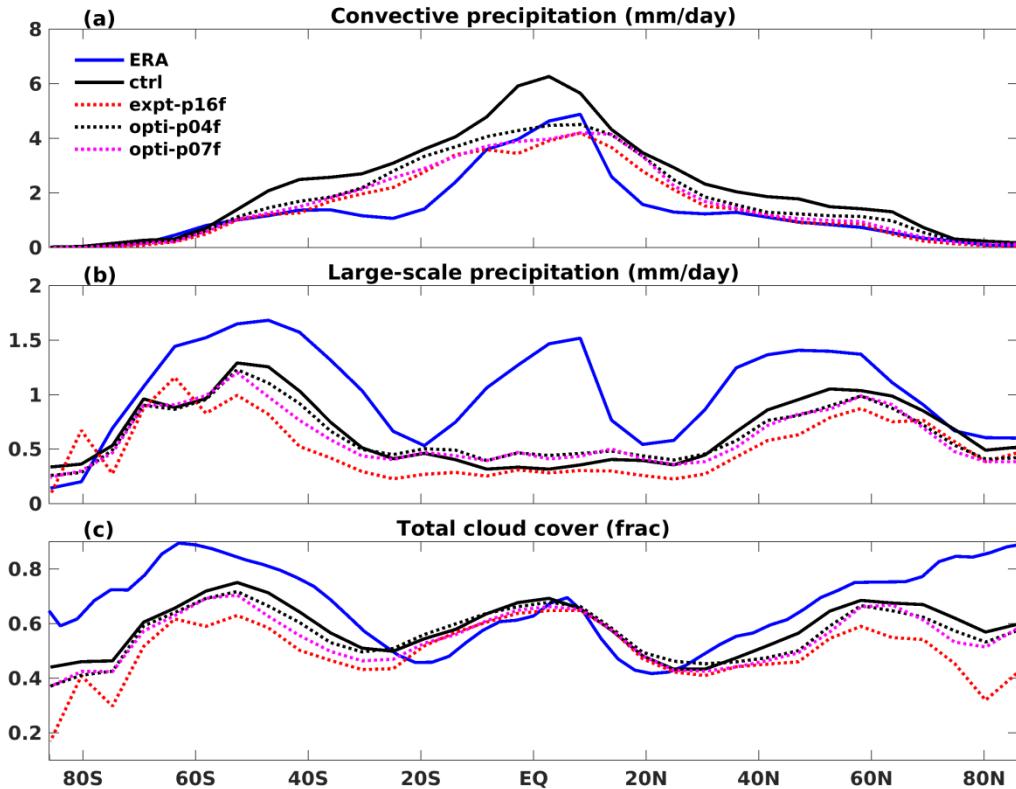


Figure 7.36. (a) Zonally averaged convective precipitation, (b) large-scale precipitation and (c) total cloud cover for the ERA-Interim data, the control run and the three simulations with optimized parameters.

The changes of precipitation and total cloud cover are also evaluated. As Figure 7.36 (a) shows, the control simulation simulates too much convective precipitation due to strong cumulus convection in the tropics. All the three simulations with optimized parameters reduce the convective precipitation by around 20% because the improved temperature caused by changes of parameter  $th2oc$  (see Figure 7.33 b-d) reduces the cumulus convection and therefore reduces the convective precipitation. There is still too much convective precipitation around 20S and 20N which is likely due to the coarse resolution of the model. In the tropics, both the large-scale precipitation and total cloud cover in Expt-p04f and Expt-p07f are slightly improved. However, they are degraded beyond the tropics. Although expt-p17f significantly degrades the SW and LW fluxes, it still slightly improves convective precipitation and total cloud cover in the tropics. This is because that the improvements in the

tropics are mainly caused by temperature improvements, due to changes of parameter *th2oc*. While beyond tropics, expt-p17f performs worse than all other simulations.

With the optimized parameters, the moisture still cannot be improved, especially in the tropics in model level 8. Another assimilation experiment (not shown here) is performed, which only includes the mean specific humidity in model level 8 in the cost function and uses 16 control parameters as in the first group assimilation experiments. The specific humidity cost cannot be reduced. It is likely that the moisture bias is due to model deficiency and cannot be reduced by tuning process parameters. With less moisture in the atmosphere, the large-scale precipitation and total cloud cover also cannot be further improved.

## 7.5 Summary

In this section, we further applied the modified DSPE method to assimilate the ERA-Interim data. Temperature, vorticity, divergence and specific humidity from the 6-hourly ERA-Interim data were used as nudging variables to synchronize chaotic behavior of the PlaSim simulation and the ERA-Interim data. Two groups of assimilation experiments were performed to investigate the performance of the method.

The first group of assimilation experiments consisted six experiments with different nudging timescales ranging from  $\tau=2.03$  days to  $\tau=0.09$  days. The cost function included annual mean vorticity, divergence, temperature and surface pressure and 16 parameters from different parameterizations were chosen as control parameters. In the cost function, the temperature cost reduction contributes most to the whole cost function reduction. The divergence and vorticity observations seem too noisy to be used for the optimization. After the optimization, the divergence cost is also reduced due to the improvement of the air temperature. In all the 16 parameters, only parameters *tfrc2* were very well informed by temperature observations and were robustness estimated, independent of the nudging timescales. Although parameters *th2oc* and *acllwr* also contributed significantly to the cost reduction, they were over-tuned without net LW flux at the surface and at the TOA. As for the nudging timescales, a small nudging timescale such as  $\tau=0.09$  days is not recommended because: (1) a small nudging timescale may seriously degrade the diagnostic variables such as cloud cover as Figure 7.13 displays and therefore influenced the estimation of parameters related to these variables; (2) it also increased the iteration number of the minimization and therefore increased computational costs. For the “maximal” configuration, a nudging timescale of  $\tau=2.03$  days worked well.

In the second group of assimilation experiments, the same observations were used for synchronization while observations included in the cost function and the control parameters were different. Two assimilation experiments were performed. Expt-p04 attempted to optimize four control parameters and to reduce bias of temperature, net surface LW flux, and net top LW flux. Expt-p07 further included net surface SW and net top SW flux in the cost function and attempted to reduce the biases by optimizing three more parameters related to cloud optical properties. The results show that the biases of net LW flux at the surface and at the TOA can be significantly reduced by optimizing parameters *th2oc* and *acllwr*. However, Expt-p04 slightly reduces the global mean net surface SW flux bias by increasing the total

cloud cover in the tropics. As the results of Expt-p07 show, the global mean net surface SW fluxes bias could be efficiently reduced by tuning parameters *tswr1* and *twsr2*. However, changes of parameters *tswr1* and *tswr2* also strongly degrade net SW flux at the TOA. The negative effects on net top SW flux are compensated by changes of parameter *tswr3*. For the air temperature, parameter *th2oc* still contribute most to reduce the temperature bias. Parameters *tfrc2* and *vdiff\_d* also contribute considerably to improve the air temperature. Overall, the bias global mean LW flux at the surface and at the TOA could be reduced by tuning parameters which controls absorptivity of clouds and water vapor and the bias global mean SW flux at the surface and at the TOA could be efficiently reduced by tuning parameters which controls optical properties of clouds. The results of the two group of assimilation experiments indicate that the net LW flux and net SW flux at the surface and at the TOA are important observations and must be included in the cost function when estimating parameters related to clouds optical properties and absorptivity coefficients of clouds and water vapor.

At last, three free model (without the nudging terms) simulations with the estimated parameters from assimilation experiments expt-p16 (with a nudging timescale of  $\tau=2.03$  days), Expt-p04 and Expt-p07 were performed to evaluate whether the estimated parameters improved the free model's climatology. All the three simulations reduce the temperature RMSE, especially in the southern hemisphere and in the tropics. In the northern hemisphere, the estimated parameters have both positive and negative effects. For net LW flux at the surface and at the TOA, both Expt-p07f and Expt-p04f reduce the global mean bias as in the corresponding assimilation experiment Expt-p07 and Expt-p04. With three more estimated parameters which controls cloud optical properties, Expt-p07f also improves SW flux at the surface and at the TOA. The results indicate that the parameters estimated based on the slave system improve the free model's climatology. For other variables that were not explicitly included in the cost function, the optimized parameters had both positive and negative effects depending on geographical regions. The optimized parameter *tfrc2*, which is informed by the temperature observations in the assimilation experiments, reduces the strength of westerlies while also reduces the variance. In the tropics, the control run simulated too much convective precipitation, and it is reduced by more than 20% because of the improved temperature fields. Both the large-scale precipitation and cloud cover are slightly improved in the tropics while are degraded at middle and high latitudes. For the specific humidity, the model simulations with optimized parameters still simulate too less moisture. The bias of moisture in the model simulation seems cannot be improved by tuning the process parameters alone. The bias of moisture is likely due to model deficiency such as coarse resolution or model structure (spectral transform method). Due to the less moisture, the total cloud cover beyond the tropics and the large-scale precipitation is less than the ERA-Interim data and by now cannot be improved by tuning process parameters.

Finally, let us answer the question we posed: (1) Using all the prognostic variables in the cost function is not helpful, and only temperature provides useful information for the optimization. Also, not all parameters can efficiently change the model's climatology. In PlaSim, only parameters *acllwr* and *th2oc* which tune absorption coefficients of clouds and water vapor seems significantly influence the air temperature. Parameter *tfrc2* also can

improve the upper-level temperature through thermal wind relation; (2) It is important to include net LW flux and net SW flux at the surface and at the TOA in the cost function. The global mean bias of SW and LW fluxes can be significantly reduced by tuning several parameters which tune properties of clouds and water vapor. However, the local errors at different latitudes or at each latitude-longitude grid point cannot be efficiently reduced, especially for the net SW flux at the surface and at the TOA; (3) Similar to the assimilation experiments, the free model simulations with optimized parameters reduce the biases, especially the radiative fluxes. However, the specific humidity cannot be improved by tuning the process parameters. The moisture bias is probably due to model deficiency. Because of this, the cloud cover and large-scale precipitation also cannot be efficiently improved beyond the tropics.

# Chapter 08

## Conclusions and Future works

Climate models use parameterizations to represent the impact of unresolved physical processes on the large-scale flow regarding large-scale flow itself. Constant parameters are unavoidably introduced in the parameterizations and values of these parameters cannot be directly observed. These parameters are an important source of model uncertainty and should be tuned to improve the model simulation. In this study, we applied the adjoint method to calibrate a climate model by optimizing process parameters. To overcome the limited assimilation window problem in the adjoint method, synchronization, which is implemented as nudging technique, was exploited to regularize nonlinear instability. For parameter estimation, constant nudging coefficients were used, and the nudging strength should at least be strong enough to push the positive Lyapunov exponents to negative values. In this method, the feasible assimilation window can be extended the arbitrary which enables parameter estimation on climate timescales.

The following part of this thesis will be separate into two parts. The first part will give main results and conclusions of the method. The second part will outline the possible future works.

### 8.1 Overview of Conclusions

In the adjoint method, the feasible assimilation window is limited by the predictability of the chaotic system. To estimate uncertain parameters and unobserved model states from a small subset of model state and then make predictions, Abarbanel et al. (2010) purposed to use chaos synchronization to extend the feasible assimilation and make full use of the observations distributed over a long period. They advocated that the nudging terms can be penalized in the cost function and break down to 0 after the optimization.

We found that the penalty terms for nudging coefficients influence the successfulness of this method, especially with a very large assimilation window. Based on Lorenz (1963) model, we showed that: a large penalty term leads to fast decay of the nudging coefficient until chaos reappears, which stalls the optimization; small or non-existing penalty terms leads to increasing nudging coefficients with noisy observations. With constant nudging coefficients, we found that parameter errors information can still be observed in the cost function and their sensitivities can be very well represented by the adjoint model. Therefore for parameter estimation, we used constant nudging coefficients to synchronize the observations and the model. Performance of this method depends on synchronization efficiency which is influenced by observation noise, observation frequency, variables chosen for nudging and nudging coefficients. With noisy and sparse observations, optimal nudging coefficients which best recovers true signal can be predefined and benefit the parameter estimation.

This method was then applied to PlaSim (Fraedrich et al., 2005b). The performance of this method was firstly validated with identical twin experiments and then applied to

assimilate the ERA-Interim data. Two configurations were used which are different in nonlinearity. For the “minimal” configuration which excludes the strong nonlinear moisture parameterizations, the feasible assimilation window is limited to around 20 days. A nudging timescale  $\tau=2.0$  days was predefined which stops the exponential error growth of TLM. Through identical twin experiments with an assimilation window of 1 year, it was seen that the default values of process parameters were accurately and efficiently retrieved within around 30 iterations, which is much more efficient than the SPSA method (Agarwal, 2016).

For the “maximal” configuration, including the moisture parameterizations significantly increased the nonlinearity and degraded the usefulness of the TLM and adjoint model. This system can be synchronized with nudging timescales smaller than 2.03 days by nudging divergence, vorticity, moisture, and temperature. The usefulness of the adjoint model was closely examined. Identical twin experiments show that this method accurately and efficiently retrieved the default values of parameters.

We then applied this method to assimilate the ERA-Interim data with the ‘maximal’ configuration. Two groups of assimilation experiments were performed with different control parameters and different observations in the cost function.

In the first group assimilation experiments, it was seen that the temperature observations are most important observations and inform several control parameters such as *th2oc*, *tfrc2*, and *acllwr*. The divergence observations and the vorticity observations are too noisy to be used for optimizing the parameters. Among all the 16 parameters, only *th2oc*, *tfrc2*, and *acllwr* are informed by the temperature observations and converge to similar values in all experiments while others diverge. The divergence of parameters can have three reasons: (a) Some sensitive observations such as SW flux and LW flux at the surface and at the TOA are not included in the cost function, and therefore the parameters related to cloud optical properties may not be very well informed. (b) Some parameters such as *vdiff\_lamm* have regional different impacts on the state variables (especially the divergence and the vorticity) which can be positive in some places and negative in others. Changing the nudging timescales is found to easily impact the pattern of the remaining model-observation misfits. This leads to a changing error that is projected on the parameter estimates and therefore to changing estimated depending on the nudging timescales. (c) Small nudging timescales significantly degrade the cloud cover fields as Figure 7.13 displays and therefore influence the estimates of parameters controlling clouds properties. In our model, it was seen that the air temperature is more sensitivity to parameters tuning absorption coefficients (longwave) of water vapor and clouds.

In the second group assimilation experiments, the bias of net LW flux at the surface is efficiently reduced by decreasing parameter *th2oc*, which reduces the absorptivity of water vapor to longwave radiation. The bias of net LW flux at the TOA is reduced by changes of parameters *th2oc* and *acllwr*. The decreasing parameters *th2oc* and *acllwr* reduces absorptivity of clouds and water vapor and hence increases outgoing longwave radiation at the TOA. As for the SW flux at the TOA and at the surface, they are efficiently improved by tuning parameters *tswr1*, *tswr2* and *tswr3* which control cloud optical properties. Error compensation occurs when optimizing these three parameters. Parameters *tswr1* and *tswr2* are

tuned to reduce net SW flux to the surface by increasing albedo and backscattering coefficients of clouds. However, they also significantly increase reflected SW flux to the TOA and significantly degrades net SW flux at the TOA. Parameter *tswr3* are tuned to reduce the negative effects on net SW flux at the TOA by reducing single scattering albedo. For the air temperature, parameter *th2oc* contributes most. Parameter *tfrc2* is very well informed by the temperature observations and improves the air temperature in upper levels. The results demonstrate that the radiative flux observations must be included when optimizing parameters controlling clouds and water vapor properties. The air temperature observations also play an important role in informing the parameters. By optimizing several process parameters controlling properties of clouds and water vapor, the biases of global mean net SW flux and net LW flux at the surface and at the TOA can be significantly improved. However, there are still errors depending latitudes which can hardly be reduced by tuning the parameters alone.

At last, effects of the estimated parameters on the free model's climatology (without nudging terms) were evaluated. The improvements of the free model's climatology resemble the improvements in the corresponding assimilation experiments. The free model simulation Expt-p04f reduces the global mean net LW flux at the surface and at the TOA by around 32 Wm<sup>-2</sup> and 12 Wm<sup>-2</sup> and matches reasonable well with the ERA-Interim data. However, the net SW flux at the surface and at the TOA is seldom changed. Expt-p07f best simulates net SW flux and net LW flux at the surface and at the TOA. While expt-p16f significantly degrades radiative fluxes because the control parameters in the assimilation experiment expt-p16 are tuned to reduce temperature bias without regarding radiative fluxes. The temperature RMSE in all the three simulation is reduced by more than 10% depending on model levels and geographical regions. In the three free-model simulations, the too strong westerlies are also significantly reduced by changes of parameter *tfrc2*. The convective precipitation is also reduced by more than 20% in the tropics. The large-scale precipitation and the total cloud cover show both improvement and degradation. However, the three simulations cannot improve the specific humidity which is likely caused by model deficiency.

Overall, by using nudging terms to synchronize the model simulation and the observations, the adjoint method is successfully applied to estimate process parameters on climate timescales and the estimated parameters improve the free model simulation. An advantage of the adjoint method is that it is much more efficient than the cost function based gradient approximation method such as SPSA (Spall, 1998). Agarwal (2016) applied the SPSA method to PlaSim for parameter estimation and found that the SPSA method needs at least 300 iterations to converge while the adjoint method usually converges within 40 iterations (one iteration costs about 3-4 times of one forward integration). The limitation of the adjoint method is that the usefulness of the adjoint model must be validated when including new observations. For some model variables, the TAF generated adjoint model cannot provide useful sensitivities with respect to the control parameters and therefore cannot be directly included in the cost function. For example, in the tropics, the specific humidity is dominated by cumulus convective which is implemented with multiple 'if' and 'where' statements. The adjoint model cannot provide useful gradients information of the specific humidity (on every model grid) with respect to the control parameters as Figure 6.14 and

Figure 6.15 displays. Modifying the implementation of these parameterization may help to improve the adjoint model.

## 8.2 Outlook

Further studies are still needed to:

- (1) In this study, PlaSim is an intermediate complex model with prescribed SST. Blessing et al. (2014) have developed a coupled system CEN Earth System Assimilation Model (CESAM) in which PlaSim is coupled to MITgcm (Marshall et al., 1997). The TLM and its adjoint can be automatically generated by an automatic differentiation tool TAF (Giering and Kaminski, 1998). The method will be further applied to this coupled system for testing. For the ocean model, we don't have enough data for synchronization, but a statistical adjoint method (Köhl and Willebrand, 2002; Liu et al., 2012) can be used to extend the feasible assimilation window.
- (2) A significant concern of this study is the coarse resolution of the model. A higher resolution version of CESAM is being developed with the atmospheric component has a resolution of T42 with ten levels, and the ocean component has a resolution of  $1^\circ \times 1^\circ$  with 23 vertical levels. The method will be further applied to this new model version.
- (3) The CEN Earth System Assimilation Model is developed for initializing seasonal to decadal prediction using the adjoint model, which is expected to produce a dynamical balanced initial condition. By using synchronization, we will extend the assimilation window for CESAM and initialize long-term predictions using the adjoint method. Moreover, parameters controlling the surface heat flux and momentum flux will be expanded geographic-dependent and time-dependent (Liu et al., 2012; Sugiura et al., 2008) and be estimated using the adjoint method to improve the model performance.

## Figure Captions

Figure 2.1. A prior probability distribution of the uncertain parameter $m$ (blue solid line), a likelihood function of observing the observations $d_o$ given the uncertain parameter $m$ (red solid line) and the posterior probability distribution estimated based on Bayes' theorem (black solid line). The dash lines represent the maximum probabilities of the three probability density function. ....	6
Figure 3.1. A schematic diagram of the tangent linear model. The black line represents a reference trajectory $X^r(r)$ . The blue line represents a perturbed trajectory $X^p(t)$ . The red stair-like line represents a tangent linear approximation based on the reference trajectory. The symbols $M_{0 \rightarrow t}(X_0)$ and $M_{0 \rightarrow r}(X_0)$ represent a model forecast operator and the tangent linear model, respectively. ....	13
Figure 3.2. A schematic diagram of the recursive minimization process in the adjoint method. ....	15
Figure 3.3. The cost function for an integration time of 5 TUs (black solid line) and 2 TUs (black dashed line). The red markers represent cases which reach the global minimum after the optimization and the black markers represent cases which are trapped into one of the secondary minima. The vertical lines indicate the minima after the optimization. ....	17
Figure 3.4. (a) The instantaneous error growth with small perturbations on parameter $\rho$ : the red line is computed based on a reference integration and a perturbed integration of the nonlinear model; the black line is from the tangent linear model; (b) is similar to (a) but the perturbations are on the initial state. The black dashed line is a linear fit to the tangent linear error growth. The perturbation size is $10^{-5}$ . ....	18
Figure 3.5. Two trajectories starting with the same initial state but slight different values of parameter $\rho$ : $\rho=28.0$ (the blue line) and $\rho=28.00001$ (the black dotted line). ....	19
Figure 3.6. A schematic diagram of synchronizing two chaotic systems. System 1 and system 2 are coupled through $(X_1, Y_1)$ and $(X, Y)$ . ....	20
Figure 3.7. The instantaneous local error growth for perturbed initial state (b) and perturbed parameter $\rho$ (a) with perturbation sizes of $10^{-5}$ for the synchronized system (3.19)-(3.21): the black dashed lines are from the tangent linear model and the red dashed lines are from perturbed the nonlinear system integrations. The nudging coefficient is $\alpha=20$ . ....	22
Figure 3.8. The cost function depending on nudging coefficient $\alpha$ and parameter $\rho$ without the penalty term (a) and with the penalty term and $W_\alpha=1.0$ (b). The black rectangle area of (b) is zoomed in as (c). The estimated nudging coefficient $\alpha$ and parameter $\rho$ for expt1 and expt2 (in Chapter 4) are marked with the green and red filled circles. The arrows represent schematic optimization routes for different experiments. The integration time is 20 TUs. ....	23
Figure 4.1. (a) norm of $\delta\rho$ , (b) the nudging coefficient $\alpha$ , (c) the cost function components $J_p$ and $J_e$ , and (d) norm of gradient , as function of iteration number for perfect observations. The line colors represent different experiments indicated by legend in (a). The dashed lines and solid lines in (c) represent $J_p$ and $J_e$ component of the cost function as in equation (3.22). ....	26

Figure 4.2. (a) norm of $\delta\rho$ , (b) the nudging coefficient $\alpha$ , (c) the cost function components $J_p$ and $J_e$ , and (d) norm of gradient , depending on iteration numbers. The line colors represent different experiments indicated by legend in (a). The dashed lines and solid lines in (c) represent $J_p$ and $J_e$ component of the cost function as in equation (3.22).....	27
Figure 4.3. (a) The cost function depending on the parameter $\rho$ and the nudging strength $\alpha$ for the case of nudging $y$ . The thick black lines represent global minima for a specific nudging strength $\alpha$ and the black circles are the estimated parameter $\rho$ from the optimization. (b) Root mean square errors compute from error-free observations with the slave system with true parameter (black circles) and optimized parameter (black line)..	29
Figure 4.4. (a) The estimated parameter $\rho$ depending on assimilation window for nudging $y$ . Expt5 (blue line) uses a nudging coefficient $\alpha=15$ with noisy $y$ observations. Expt6 (black line) uses a nudging coefficient $\alpha=60$ with noisy $y$ observations and Expt7 (red line) uses a nudging strength $\alpha=60$ with error-free $y$ observations. (b) Root mean square errors for the above experiments. The dashed line represents the noise error introduced by nudging term shown in Figure 4.3(b). .....	31
Figure 4.5. Root mean square errors computed from error-free observations and the slave system with different sets of variables chosen for nudging (see legend). The red markers represent the optimal nudging coefficient which best recovers the error-free signal from noisy observations. The dashed line is noise level. ....	32
Figure 4.6. The estimated parameter $\rho$ depending on the assimilation window with different sets of variables chosen for nudging (see legend). ....	32
Figure 4.7. Cost functions depending on the parameter $\rho$ and the nudging strength $\alpha$ with nudging $y$ for an observation frequency of every 5 steps, 10 steps and 20 steps as indicated by the titles. The thick black lines represent the global minimum with different nudging coefficients, the black triangles are the estimated parameter $\rho$ after the optimization. ....	33
Figure 4.8. Trajectories of the variable $y$ . The blue line is the true trajectory; the dotted lines are the linear interpolated trajectories with observations every 5steps (red dots) and 20 steps (magenta dots). The solid lines are trajectories after optimization with the optimal nudging strength for observations every 5steps (red line)and 20 steps (magenta line). ...	34
Figure 4.9. The estimated parameter $\rho$ depending on the assimilation window for nudging $y$ with the optimal nudging strength for different observation frequencies (see legend). ...	35
Figure 5.1. Annual mean of zonal averaged zonal wind (m/s) and temperature ( $^{\circ}$ C) for the ERA-Interim reanalysis (a) and the PlaSim simulation (b). Annual mean of the ERA-Interim data use data of year 1992.....	39
Figure 5.2. Zonal averaged climatological temperature difference ( $^{\circ}$ C) (a) and specific humidity difference (kg/kg) (b) between the PlaSim simulation and the ERA-Interim reanalysis. The difference is computed by the PlaSim simulation minus ERA-Interim reanalysis. Annual mean of ERA-Interim data uses data of year 1992. ....	40
Figure 5.3. Temperature difference in level 10 between the PlaSim simulation and the ERA-Interim reanalysis for model level 10. The difference is computed by the PlaSim simulation minus ERA-Interim data. Annual mean of the ERA-Interim data uses data of year 1992. The contour interval is 2 $^{\circ}$ C. ....	40

Figure 6.1. The time evolution of RMSEs for the exact TLM (the green line), the modified TLM (the black line) and a perturbed PlaSim integration (the blue line) with a perturbation size of 0.0001%. The red dashed line is a linear fit of the blue line for the first 50 days.....	46
Figure 6.2. Spatial distribution of the surface pressure difference (10 Pa) caused by 10% perturbations on all 9 parameters: (a) the finite difference between a perturbed PlaSim integration and the control integration, (b) the modified TLM, (c) the exact TLM. The integration time is 10 days.....	48
Figure 6.3. The cost function depending on parameter perturbations (%) for an assimilation window of 10 days. All 9 parameters are perturbed. ....	48
Figure 6.4. Convergence of the minimization for the minimal configuration for four experiments with different perturbation size (see legend) on all 9 parameters. The assimilation window is 10 days. ....	49
Figure 6.5. The time evolution of instantaneous RMSEs for the modified TLM without nudging ( $\tau=0$ day), with nudging timescales ranging from 5 days to 2 days (see legend). ....	50
Figure 6.6. Cost function components depending on the size of parameter perturbations for PlaSim (the solid lines) and the slave system (the dashed lines): (a) surface pressure, (b) temperature, (c) vorticity and (d) divergence. The integration time is 1 year. The cost function is normalized by the mean value over all perturbations. ....	51
Figure 6.7. The surface pressure difference (Pa) caused by a 10% perturbation on all 9 parameters for the finite difference of perturbed slave system integrations (a) and the TLM of the slave system (b) after 1 year integrations.....	52
Figure 6.8. Convergence of the minimization for 3 assimilating experiments with an assimilation window of 1 year: (a) 10% perturbations (b) 20% perturbations (c) -90% perturbations. The cost function, norm of gradients and norm of parameter perturbation shown in the figure are normalized by their maximum values, respectively.....	52
Figure 6.9. Temperature difference ( $^{\circ}$ C) of level 01(uppermost) caused by a 10% perturbation on <i>tfrc1</i> after a 2-days integration of the nonlinear model (a) and the TLM (b). Temperature difference ( $^{\circ}$ C) of level 10 (lowermost) caused by a 10% perturbation on <i>vdiff_lamm</i> after a 6-hours integration of PlaSim (c) and the TLM (d).....	54
Figure 6.10. The time evolution of instantaneous RMSEs for the TLM without nudging ( $\tau=0$ day), with nudging timescales ranging from 6.2 days to 0.25 days (see legend). The “maximal” configuration is used. ....	55
Figure 6.11. The cost functions depending on sizes of perturbations on the 13 parameters (see legend). The integration time is 2 months. ....	56
Figure 6.12. The zonal mean temperature changes from the TLM (b and d) and the finite differences of perturbed slave system integrations (a and c) with a -10% perturbation on <i>tfrc1</i> and <i>th2oc</i> : (a) and (b) are caused by perturbed <i>tfrc1</i> , (b) and (d) are caused by perturbed <i>th2oc</i> . The integration time is 1 year. ....	57
Figure 6.13. The temperature changes of level 10 for the TLM (b) and the finite difference of perturbed slave system integrations with a -10% perturbation on the parameter <i>th2oc</i> ..	57
Figure 6.14. The zonal mean specific humidity (kg/kg) changes from the finite differences of perturbed slave system integrations (a and c) and the TLM (b and d) with a 10%	

perturbation on <i>tfrc1</i> and <i>th2oc</i> : (a) and (b) are caused by perturbed <i>tfrc1</i> , (b) and (d) are caused by perturbed <i>th2oc</i> . The integration time is 1 year.....	58
Figure 6.15. Specific humidity (kg/kg) changes in level 10 for the TLM (b) and the finite difference of perturbed slave system integrations with a -10% perturbation on the parameter <i>th2oc</i> . The integration time is 12 months. The scales of color bars are different for (a) and (b). .....	59
Figure 6.16. The cost function components of specific humidity depending on parameter perturbations for 13 parameters (see legend).....	59
Figure 6.17. Convergence of the minimization for 4 experiments with perturbation sizes of 10%, -10%, 20% and -20% (see legend): (a) the cost function, (b) norm of the perturbation sizes, (c) norm of the gradients. ....	60
Figure 6.18. Absolute difference of components of control parameters over iteration number. ....	61
Figure 6.19. As Figure 6.18 but for gradients of the cost function with respect to the control parameters.....	62
Figure 6.20. The new search directions (d), the diagonal part (d_diag) and off-diagonal (d_offdiag) and the gradients information from the adjoint model (g) for the parameter <i>tdissq.in</i> experiments with -20% perturbations. The components are computed based on equations (3.20)-(3.23).....	63
Figure 7.1. Ratios of the nudging terms to the total tendencies for temperature (a), moisture (b), divergence (c) and vorticity (d). The ratios are averaged over a 6-month integration period.....	66
Figure 7.2. Convergence of the minimization for the cost function and the gradient norm in all six experiments.....	67
Figure 7.3. Convergence of all 16 parameters (see title) in the 6 experiments.....	68
Figure 7.4. Constituents of the cost function reduction in the 6 assimilation experiments. ....	69
Figure 7.5. (a) Zonally averaged temperature difference (°C) of the slave system simulation and the ERA-Interim data. (b) Zonally averaged temperature increment after the optimization in the assimilation experiment with a nudging timescale of $\tau=2.03$ days. The contour interval is 0.5 °C.....	70
Figure 7.6. (a) The temperature difference (°C) in level 10 between the slave system simulation and the ERA-Interim data. (b) The temperature increment (°C) in level 10 after the optimization of the slave system. The nudging timescale of the slave system is $\tau=2.03$ days. The contour interval is 1 °C.....	70
Figure 7.7. The divergence differences between the slave systems and the ERA-Interim data with nudging timescales of $\tau=2.03$ days (a) and $\tau=0.09$ days (c) in model level 10. The divergence increments after the optimization are shown in (b) and (d), respectively. ....	71
Figure 7.8. Zonally averaged differences and zonally averaged increments of temperature (a), divergence (b), vorticity (c) and surface pressure (d) in model level 10. The solid lines represent the differences of the slave system with different nudging timescales and the dashed lines are the increments, respectively. The colors represent the assimilating experiments with different nudging timescales as the legend indicates. ....	72

Figure 7.9. Contribution of each parameter (see legend) to constituents of the total cost function reduction in all 6 experiments (see titles).....	73
Figure 7.10. Zonally averaged temperature increments ( $^{\circ}\text{C}$ ) caused by changes of parameter $th2oc$ in the assimilation experiments with nudging timescales of $\tau=2.03\text{days}$ (a) and $\tau=0.09\text{days}$ (b). The contour interval is $0.5\text{ }^{\circ}\text{C}$ in (a) and $0.05\text{ }^{\circ}\text{C}$ in (b).....	74
Figure 7.11. Divergence increments ( $1/\text{s}$ ) caused by changes of parameter $th2oc$ in the assimilation experiments with nudging timescales of $\tau=2.03\text{days}$ and $\tau=0.09\text{days}$ .....	75
Figure 7.12. Zonally averaged temperature increments ( $^{\circ}\text{C}$ ) caused by changes of $acllwr$ in the assimilating experiments with nudging timescales of $\tau=2.03\text{days}$ and $\tau=0.09\text{days}$ . The contour interval is $0.25\text{ }^{\circ}\text{C}$ in (a) and $0.01\text{ }^{\circ}\text{C}$ in (b).....	76
Figure 7.13. Zonally averaged cloud cover in the control run and with nudging timescales of $\tau=2.03\text{days}$ and $\tau=0.09\text{days}$ .....	76
Figure 7.14. Zonally averaged temperature increments ( $^{\circ}\text{C}$ ) caused by changes of parameter $tpofmt$ in the assimilation experiments with nudging timescales of $\tau=2.03\text{days}$ and $\tau=0.09\text{days}$ . The contour interval is $0.01\text{ }^{\circ}\text{C}$ in (a) and $0.02\text{ }^{\circ}\text{C}$ in (b).....	77
Figure 7.15. Zonally averaged temperature increments ( $^{\circ}\text{C}$ ) caused by changes of parameter $tfrc2$ in the assimilating experiments with nudging timescales of $\tau=2.03\text{days}$ (a) and $\tau=0.09\text{days}$ (b). (c) and (d) are the same as (a) and (b) except that (c) and (d) are the divergence increments.....	77
Figure 7.16. (a) Zonally averaged temperature differences (solid lines) and zonally averaged temperature increments (dashed lines) in level 04 in all the assimilating experiments. (b) Similar to (a) but for divergence and in level 02. The line colors represent the assimilation experiments with different nudging timescales as the legend indicates.....	78
Figure 7.17. Divergence difference (a) and divergence increment (b) in level 3 caused by changes of parameter $tdissd$ in the assimilation experiment with a nudging timescale of $\tau=2.03\text{days}$ .....	79
Figure 7.18. Divergence differences (solid lines) and divergence increments (dashed lines) along 20S in level 3 in all the assimilating experiments. The line colors represent the assimilating experiments with different nudging timescales as the legend indicates.....	80
Figure 7.19. Contribution of parameter $vdiff\_lamm$ to the divergence cost reduction and the vorticity cost reduction in the assimilation experiments with different nudging timescales (horizontal axis). The bar colors indicate model levels as the legend shows.....	80
Figure 7.20. (a) Divergence differences (solid lines) and divergence increments (dashed lines) along 20S in level 3 in all the assimilating experiments. (b) Same as (a) but for vorticity. The line colors represent the assimilation experiments with different nudging timescales as the legend indicates.....	81
Figure 7.21. (a) Zonally averaged temperature differences (solid lines) and temperature increments (dashed lines) in level 10 caused by changes of parameter $vdiff_d$ in all assimilating experiments. (b) Same as (a) but for divergence. The line colors represent the assimilation experiments with different nudging timescales as the legend indicates.	82
Figure 7.22. Temperature increments ( $^{\circ}\text{C}$ ) by changes of parameters $tswr1$ (a), $tswr2$ (b) and $tswr3$ (c) in the assimilation experiment with a nudging timescale of $\tau=2.03\text{days}$ .....	83

Figure 7.23. Convergence of the cost function (a), parameters <i>tfr2</i> (b), <i>th2oc</i> and <i>acllwr</i> (c) in an assimilation experiment with only temperature observations in the cost function. The nudging timescale is $\tau=2.03\text{days}$ .	83
Figure 7.24. Convergence of the total cost (a), the temperature cost (b), sum of the net SW flux costs at the surface and at the TOA (c) and sum of the net LW flux costs at the surface and at the TOA vs. iteration number. The solid lines with stars are from Expt-p04 while the dashed lines with circles are from Expt-p07. In Expt-p04, net LW flux at the surface and at the TOA are not included in the cost function and therefore (d) only shows sum of the net LW flux costs at the surface and at the TOA from Expt-p07.	85
Figure 7.25. Temperature increments in both the assimilation experiments.	87
Figure 7.26. Approximated contributions of each parameter to the temperature cost reduction and the surface pressure cost reduction in the two assimilation experiments. The blue bars are the temperature cost reduction and the yellow bars are the surface pressure cost reduction. Positive values (negative values) mean the control parameters reduced (increased) the cost by the values on y-axis.	87
Figure 7.27. Zonally averaged net surface LW flux (a), net top LW flux (b), net top SW flux (c) and net surface SW flux (d) in the ERA-Interim data (blue lines), the slave system simulation (black lines) and the two assimilation experiments (dotted lines with markers, see legend). Negative (positive) values mean upward (downward).	89
Figure 7.28. Zonally averaged totally cloud cover in the slave system simulation (black line), the ERA-Interim data (blue line) and in the two assimilation experiments (dotted lines with markers, see legend).	89
Figure 7.29. Approximated contributions of each parameter (see legend in d) to net surface LW flux (a), net top LW flux (b), net top SW flux (c) and net surface SW flux (d). Positive values (negative values) mean the control parameter increases (decreases) the corresponding variables. The corresponding values of ERA-Interim data and the slave system simulated values are on top of each subplot.	91
Figure 7.30. Increments of net surface SW flux (a)-(c) and net top SW flux (d)-(f) caused by changes of parameters <i>tswr1</i> (a,d), <i>tswr2</i> (b,e) and <i>twsr3</i> (c,f).	92
Figure 7.31. Meridional distributions of the annual mean difference of surface LW flux (a), top LW flux (b), surface SW flux (c), top SW flux (d), surface latent heat flux (e) and surface sensible heat flux between the model simulations and the ERA-Interim data. The line style and line color represent the model simulations with optimized parameters from different assimilation experiments (see legend).	95
Figure 7.32. The root mean square error of temperature (upper panels) and Zonal wind (lower panels) for different geographical regions. The RMSE is computed with the 4 model simulations (see legend) and the ERA-Interim data.	96
Figure 7.33. (a) Annual mean of zonally averaged temperature difference between the control run and the ERA-Interim data. Annual mean of zonally averaged temperature differences between the model simulations with optimized parameters from different assimilation experiments and the control run (b-d).	97
Figure 7.34. (a) Annual mean of zonally Zonal wind for the ERA-Interim data(a), the control simulation(b) and Expt-p07f(c).	98

Figure 7.35. (a) Specific humidity difference between the control simulation and the ERA-Interim data, (b) specific humidity difference between Expt-p07f and the control simulation. The contour interval for the dashed and solid lines is $0.5 \times 10^{-3}$ kg/kg. ....	98
Figure 7.36. (a) Zonally averaged convective precipitation, (b) large-scale precipitation and (c) total cloud cover for the ERA-Interim data, the control run and the three simulations with optimized parameters. ....	99

## Table Captions

Table 4.1. Parameter $\rho$ , initial state and the nudging coefficient $\alpha$ after the optimization for case studies.....	25
Table 5.1. Process parameters used in parameter estimation .....	41
Table 7.1 Experiment setups for the second groups of assimilation experiments. ....	67
Table 7.2. The default values and the estimated values of the parameters after the optimization. ....	85
Table 7.3. The net SW flux and net LW flux at the surface and at the TOA in the ERA-Interim data, the slave system simulation and the assimilation experiments. ....	88
Table 7.4. Radiative fluxes, surface latent heat flux and surface sensible heat flux in the ERA-Interim data and the model simulations with parameters from different assimilation experiments. Negative values represent outgoing. ....	94

## Reference

- Abarbanel, H. D., M. Kostuk, and W. Whartenby, 2010: Data assimilation with regularized nonlinear instabilities. *Quarterly Journal of the Royal Meteorological Society*, **136**, 769-783.
- Abarbanel, H. D., D. R. Creveling, R. Farsian, and M. Kostuk, 2009: Dynamical state and parameter estimation. *SIAM Journal on Applied Dynamical Systems*, **8**, 1341-1381.
- Alvarez, J., 1996: Synchronization in the Lorenz system: stability and robustness. *Nonlinear Dynamics*, **10**, 89-103.
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Monthly weather review*, **129**, 2884-2903.
- Annan, J., J. Hargreaves, N. Edwards, and R. Marsh, 2005: Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean modelling*, **8**, 135-154.
- Bao, S. a., 1993: Optimal determination of nudging coefficients using the adjoint equations. *Tellus A*, **45**, 358-369.
- Blackadar, A. K., 1962: The vertical distribution of wind and turbulent exchange in a neutral atmosphere. *Journal of Geophysical Research*, **67**, 3095-3102.
- Blessing, S., and Coauthors, 2008: Adjoint Modelling for Assimilation and Diagnosis on Climate Timescales. *Dissertation*.
- Blessing, S., and Coauthors, 2014: Testing variational estimation of process parameters and initial conditions of an earth system model. *Tellus A*, **66**.
- Boccaletti, S., J. Kurths, G. Osipov, D. L. Valladares, and C. S. Zhou, 2002: The synchronization of chaotic systems. *Physics Reports*, **366**, 1-101.
- Dee, D., and Coauthors, 2011: The ERA- Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**, 553-597.
- Errico, R. M., 1997: What is an adjoint model? *Bulletin of the American Meteorological Society*, **78**, 2577-2591.
- Errico, R. M., T. Vukićević, and K. Raeder, 1993: Examination of the accuracy of a tangent linear model. *Tellus A*, **45**, 462-477.
- Evensen, G., 2009: *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media.
- Fletcher, R., and M. J. Powell, 1963: A rapidly convergent descent method for minimization. *The Computer Journal*, **6**, 163-168.
- Fraedrich, K., E. Kirk, U. Luksch, and F. Lunkeit, 2005a: The portable university model of the atmosphere (PUMA): Storm track dynamics and low-frequency variability. *Meteorologische Zeitschrift*, **14**, 735-745.
- Fraedrich, K., H. Jansen, E. Kirk, U. Luksch, and F. Lunkeit, 2005b: The Planet Simulator: Towards a user friendly model. *Meteorologische Zeitschrift*, **14**, 299-304.

- Gauthier, P., 1992: Chaos and quadri-dimensional data assimilation: a study based on the Lorenz model. *Tellus A*, **44**, 2-17.
- Giering, R., and T. Kaminski, 1998: Recipes for adjoint code construction. *ACM Transactions on Mathematical Software (TOMS)*, **24**, 437-474.
- Hoke, J. E., and R. A. Anthes, 1976: The initialization of numerical models by a dynamic-initialization technique. *Monthly Weather Review*, **104**, 1551-1556.
- Hourdin, F., and Coauthors, 2016: The art and science of climate model tuning. *Bulletin of the American Meteorological Society*.
- Järvinen, H., and Coauthors, 2010: Estimation of ECHAM5 climate model closure parameters with adaptive MCMC. *Atmospheric Chemistry and Physics*, **10**, 9993-10002.
- Jackson, C., M. K. Sen, and P. L. Stoffa, 2004: An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions. *Journal of Climate*, **17**, 2828-2841.
- Janisková, M., J.-N. Thépaut, and J.-F. Geleyn, 1999: Simplified and regular physical parameterizations for incremental four-dimensional variational assimilation. *Monthly weather review*, **127**, 26-45.
- Köhl, A., 2015: Evaluation of the GECCO2 ocean synthesis: transports of volume, heat and freshwater in the Atlantic. *Quarterly Journal of the Royal Meteorological Society*, **141**, 166-181.
- Köhl, A., and J. Willebrand, 2002: An adjoint method for the assimilation of statistical characteristics into eddy-resolving ocean models. *Tellus A*, **54**, 406-425.
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. Hnilo, M. Fiorino, and G. Potter, 2002: Ncep-doe amip-ii reanalysis (r-2). *Bulletin of the American Meteorological Society*, **83**, 1631-1643.
- Knorr, W., T. Kaminski, M. Scholze, N. Gobron, B. Pinty, R. Giering, and P. P. Mathieu, 2010: Carbon cycle data assimilation with a generic phenology model. *Journal of Geophysical Research: Biogeosciences*, **115**.
- Krishnamurti, T., J. Xue, H. Bedi, K. Ingles, and D. Oosterhof, 1991: Physical initialization for numerical weather prediction over the tropics. *Tellus A: Dynamic Meteorology and Oceanography*, **43**, 53-81.
- Kuo, H.-L., 1974: Further studies of the parameterization of the influence of cumulus convection on large-scale flow. *Journal of the Atmospheric Sciences*, **31**, 1232-1240.
- Laroche, S., M. Tanguay, and Y. Delage, 2002: Linearization of a simplified planetary boundary layer parameterization. *Monthly weather review*, **130**, 2074-2087.
- Laursen, L., and E. Eliasen, 1989: On the effects of the damping mechanisms in an atmospheric general circulation model. *Tellus A: Dynamic Meteorology and Oceanography*, **41**, 385-400.
- Le-Dimet, F. X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A*, **38**, 97-110.
- Lea, D. J., M. R. Allen, and T. W. Haine, 2000: Sensitivity analysis of the climate of a chaotic system. *Tellus A*, **52**, 523-532.
- Liu, C., A. Köhl, and D. Stammer, 2012: Adjoint-based estimation of eddy-induced tracer mixing parameters in the global ocean. *Journal of Physical Oceanography*, **42**, 1186-1206.

- Louis, J.-F., 1979: A parametric model of vertical eddy fluxes in the atmosphere. *Boundary-Layer Meteorology*, **17**, 187-202.
- Lunkeit, F., 2001: Synchronization experiments with an atmospheric global circulation model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **11**, 47-51.
- Mahfouf, J. F., 1999: Influence of physical processes on the tangent- linear approximation. *Tellus A*, **51**, 147-166.
- Marshall, J., A. Adcroft, C. Hill, L. Perelman, and C. Heisey, 1997: A finite- volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *Journal of Geophysical Research: Oceans*, **102**, 5753-5766.
- Mauritsen, T., and Coauthors, 2012: Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, **4**.
- Menemenlis, D., I. Fukumori, and T. Lee, 2005: Using Green's functions to calibrate an ocean general circulation model. *Monthly weather review*, **133**, 1224-1240.
- Miller, R. N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the atmospheric sciences*, **51**, 1037-1056.
- Mu, M., and J. Wang, 2003: A method for adjoint variational data assimilation with physical "on-off" processes. *Journal of the atmospheric sciences*, **60**, 2010.
- Murphy, J. M., D. M. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768-772.
- Navon, I. M., X. Zou, J. Derber, and J. Sela, 1992: Variational data assimilation with an adiabatic version of the NMC spectral model. *Monthly weather review*, **120**, 1433-1446.
- Nese, J. M., J. A. Dutton, and R. Wells, 1987: Calculated attractor dimensions for low-order spectral models. *Journal of the atmospheric sciences*, **44**, 1950-1972.
- Pazó, D., A. Carrassi, and J. López, 2015: Data- assimilation by delay- coordinate nudging. *Quarterly Journal of the Royal Meteorological Society*.
- Pires, C., R. Vautard, and O. Talagrand, 1996: On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus A*, **48**, 96-121.
- Polkova, I., A. Köhl, and D. Stammer, 2014: Impact of initialization procedures on the predictive skill of a coupled ocean–atmosphere model. *Climate dynamics*, **42**, 3151-3169.
- Rabier, F., H. Järvinen, E. Klinker, J. F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four- dimensional variational assimilation. I: Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, **126**, 1143-1170.
- Rui-Qiang, D., L. Jian-Ping, and H. Kyung-Ja, 2008: Nonlinear local Lyapunov exponent and quantification of local predictability. *Chinese Physics Letters*, **25**, 1919.
- Saha, S., and Coauthors, 2010: The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, **91**, 1015.
- Schiller, A., and G. B. Brassington, 2011: *Operational oceanography in the 21st century*. Springer Science & Business Media.

- Schirber, S., D. Klocke, R. Pincus, J. Quaas, and J. L. Anderson, 2013: Parameter estimation using data assimilation in an atmospheric general circulation model: From a perfect toward the real world. *Journal of Advances in Modeling Earth Systems*, **5**, 58-70.
- Slingo, A., and J. M. Slingo, 1991: Response of the National Center for Atmospheric Research Community Climate Model to improvements in the representation of clouds. *Journal of Geophysical Research: Atmospheres*, **96**, 15341-15357.
- Stammer, D., and C. Wunsch, 1996: The determination of the large- scale circulation of the Pacific Ocean from satellite altimetry using model Green's functions. *Journal of Geophysical Research: Oceans*, **101**, 18409-18432.
- Stammer, D., and Coauthors, 2002: Global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model. *Journal of Geophysical Research: Oceans*, **107**.
- Stensrud, D. J., and J.-W. Bao, 1992: Behaviors of variational and nudging assimilation techniques with a chaotic low-order model. *Monthly weather review*, **120**, 3016-3028.
- Stephens, G., 1978: Radiation profiles in extended water clouds. I: Theory. *Journal of the Atmospheric Sciences*, **35**, 2111-2122.
- Sugiura, N., and Coauthors, 2008: Development of a four- dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations. *Journal of Geophysical Research: Oceans*, **113**.
- Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quarterly Journal of the Royal Meteorological Society*, **113**, 1311-1328.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, **93**, 485-498.
- Tett, S. F., M. J. Mineter, C. Cartis, D. J. Rowlands, and P. Liu, 2013: Can top-of-atmosphere radiation measurements constrain climate predictions? Part I: Tuning. *Journal of Climate*, **26**, 9348-9366.
- Tett, S. F. B., K. Yamazaki, M. J. Mineter, C. Cartis, and N. Eizenberg, 2017: Calibrating Climate Models Using Inverse Methods: Case studies with HadAM3, HadAM3P and HadCM3. *Geosci. Model Dev. Discuss.*, **2017**, 1-36.
- Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, **117**, 1779-1800.
- Wu, X., S. Zhang, Z. Liu, A. Rosati, T. L. Delworth, and Y. Liu, 2012: Impact of geographic-dependent parameter optimization on climate estimation and prediction: simulation with an intermediate coupled model. *Monthly Weather Review*, **140**, 3956-3971.
- Xu, Q., 1996: Generalized adjoint for physical processes with parameterized discontinuities. Part I: Basic issues and heuristic examples. *Journal of the atmospheric sciences*, **53**, 1123-1142.
- Yang, S.-C., and Coauthors, 2006: Data assimilation as synchronization of truth and model: Experiments with the three-variable lorenz system\*. *Journal of the atmospheric sciences*, **63**, 2340-2354.

- Zhang, S., 2011: Impact of observation- optimized model parameters on decadal predictions: Simulation with a simple pycnocline prediction model. *Geophysical Research Letters*, **38**.
- Zhang, S., X. Zou, and J. E. Ahlquist, 2001: Examination of numerical results from tangent linear and adjoint of discontinuous nonlinear models. *Monthly weather review*, **129**, 2791-2804.
- Zhang, T., L. Li, Y. Lin, W. Xue, F. Xie, H. Xu, and X. Huang, 2015: An automatic and effective parameter optimization method for model tuning. *Geoscientific Model Development*, **8**, 3579-3591.
- Zhu, J., and M. Kamachi, 2000: The role of time step size in numerical stability of tangent linear models. *Monthly Weather Review*, **128**, 1562-1572.
- Zhu, J., H. Wang, and M. Kamachi, 2002: The improvement made by a modified TLM in 4DVAR with a geophysical boundary layer model. *Advances in Atmospheric Sciences*, **19**, 563-582.
- Zou, X., 1997: Tangent linear and adjoint of “on-off” processes and their feasibility for use in 4-dimensional variational data assimilation. *Tellus A*, **49**.
- Zou, X., I. Navon, and F. LeDimet, 1992: An optimal nudging data assimilation scheme using parameter estimation. *Quarterly Journal of the Royal Meteorological Society*, **118**, 1163-1186.
- Zou, X., I. Navon, and J. Sela, 1993: Control of gravitational oscillations in variational data assimilation. *Monthly weather review*, **121**, 272-289.
- Županski, D., and F. Mesinger, 1995: Four-dimensional variational assimilation of precipitation data. *Monthly Weather Review*, **123**, 1112-1127.
- Zupanski, M., 1993: Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Monthly Weather Review*, **121**, 2396-2408.

# **Declaration**

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

August 2, 2017

Hamburg,

Signature