

## ABSTRACT

BATHMANN, KRISTEN ASHLEY. State Estimation from Sparse Observation Networks and Satellite Measurement. (Under the direction of John Harlim.)

Filtering is the process of determining the statistically best estimate of a true signal based on partial, noisy observations. It is a two-step predictor-corrector scheme consisting of a forecast step, followed by an analysis step that updates the prediction as observations become available. In this thesis, two different filtering methods are applied under two different scenarios. In the first setting, two-dimensional irregularly spaced, sparsely observed turbulent signals are assimilated through a hierarchical Bayesian reduced stochastic filter. This approach is tested using a two-layer quasi-geostrophic model on a two-dimensional domain with a small radius of deformation to mimic ocean turbulence. The hierarchical strategy consists of two steps, combining a data-driven interpolation with the Mean Stochastic Model filter (MSM). Two interpolation schemes, a piecewise linear interpolation and ordinary kriging, are compared in this first step. Filtered estimates obtained from ordinary kriging are superior to those obtained from the linear interpolation when the observation networks are not too sparse, and the dynamical constraint of the MSM becomes important when the observation noise variance is large.

In the second half of this thesis, the potential for using cloud-impacted satellite observations to improve the analysis estimate of the Local Ensemble Transform Kalman filter is assessed. Most data assimilation schemes with satellite data utilize only clear sky infrared radiances. However, cloudy radiances contain significant information about the atmospheric state and could have a positive impact on weather prediction. This study is based on synthetic Atmospheric Infrared Sounder (AIRS) measurements, incorporating the temperature and humidity profiles of the multcloud model, a toy model that simulates tropical convection. The key parameters that induce filter bias in the presence of a cloud are identified. In particular, the cloud cover percent, and, more importantly, the cloud top pressure are the two most crucial parameters to estimate correctly. Furthermore, this study also provides error bounds on the analyses based on using cloud-impacted AIRS data. The findings here suggest that with reasonable estimates of cloud cover percent and cloud top pressure, cloudy AIRS data can be potentially as effective as clear sky data, and could enhance numerical weather prediction.

© Copyright 2014 by Kristen Ashley Bathmann

All Rights Reserved

State Estimation from Sparse Observation Networks  
and Satellite Measurement

by  
Kristen Ashley Bathmann

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

Applied Mathematics

Raleigh, North Carolina

2014

APPROVED BY:

---

Ralph Smith

---

Mark Hoefer

---

Anantha Aiyyer

---

John Harlim  
Chair of Advisory Committee

## BIOGRAPHY

Kristen Bathmann (nee Brown) grew up in Allentown, Pennsylvania. She received a Bachelor of Science in mathematics from Bucknell University in May 2009. In December 2011, she received her Master of Science in applied mathematics from North Carolina State University, where she is currently a doctoral student. She plans to graduate in 2014 with a Ph.D in applied mathematics.

## ACKNOWLEDGEMENTS

My research was partially supported with a Graduate Research Assistantship, through John Harlim's Office of Naval Research Fund N00014-11-1-0310, and by the NSF funded Research for Early Graduate Student program DMS-0943855.

I would like to thank my advisor, John Harlim, for his guidance and patience, and for his hard work in the time he spent working with me. I would also like to thank Elana Fertig, for taking the time to work with me, and for introducing me to the prototype Community Radiative Transfer Model. I also thank Emily L. Kang, who originally introduced me to kriging, and Samuel Stechmann, for sharing the multcloud model codes. Finally, I would like to thank my husband, Robert, and the rest of my family and friends, for supporting me through this process and throughout my education.

# TABLE OF CONTENTS

<b>LIST OF TABLES</b>	<b>v</b>
<b>LIST OF FIGURES</b>	<b>vi</b>
<b>Chapter 1 Introduction</b>	<b>1</b>
<b>Chapter 2 A Hierarchical Bayesian Reduced Stochastic Filter</b>	<b>5</b>
2.1 Quasi-geostrophic Model	7
2.1.1 Background Information	7
2.1.2 The Model	8
2.2 Hierarchical Bayesian Filtering Approach	15
2.2.1 Linear Interpolation	15
2.2.2 Ordinary Kriging	18
2.2.3 Reduced Stochastic Filter	24
2.3 Results	31
2.4 Summary	40
<b>Chapter 3 State Estimation from Cloudy Satellite Observations</b>	<b>42</b>
3.1 The Multicloud Model	44
3.1.1 Background Information	44
3.1.2 The Model	46
3.2 Radiative Transfer Model	51
3.2.1 Adding Cloud Cover	54
3.3 Interpolating Multicloud Model Variables to RTM Radiances	57
3.3.1 Adding Cloud Cover to Synthetic Satellite Observations	62
3.4 Data Assimilation	65
3.4.1 Local Ensemble Transform Kalman Filter	66
3.4.2 Biases from Cloudy Sky Observations	71
3.5 Results	73
3.5.1 Direct Observations	74
3.5.2 Synthetic Clear Sky AIRS Observations	76
3.5.3 Synthetic Cloudy Sky AIRS Observations	79
3.6 Summary	93
<b>References</b>	<b>95</b>

## LIST OF TABLES

Table 3.1	Parameters and values of the multcloud model . . . . .	50
Table 3.2	The values of constants used to interpolate multcloud model variables. . . .	62

## LIST OF FIGURES

Figure 2.1	The barotropic velocity field (arrows) and the streamfunction, $\psi$ , (contour) (top) and the baroclinic velocity field and streamfunctions, $\psi^c$ (bottom) at times $T = 100$ and $200$ . . . . .	14
Figure 2.2	True process, compared to a linear interpolation at $T = 340$ , with $r^o = 17.3$ . The circles give observation locations and the contours give the barotropic streamfunction, $\psi$ . . . . .	16
Figure 2.3	The noise covariance matrices associated with linear interpolation at $T = 340$ , with $r^o = 17.3$ in physical space (left) Fourier space (real components, right). The imaginary component is one magnitude order smaller (not shown). . . . .	18
Figure 2.4	An exponential variogram fit at $T = 340$ , with $r^o = 17.3$ , plotted as a function of $r$ . . . . .	20
Figure 2.5	True process, compared with kriging at $T = 340$ , with $r^o = 17.3$ . The circles give observation locations and the contours give the barotropic streamfunction, $\psi$ . . . . .	23
Figure 2.6	The noise covariance matrices associated with kriging at $T = 340$ , with $r^o = 17.3$ in physical space (left) and Fourier space (real components, right). The imaginary component is one magnitude order smaller (not shown). . . . .	24
Figure 2.7	Percentage of variances of the barotropic streamfunction as a function of modes, from largest to smallest. . . . .	28
Figure 2.8	Marginal pdfs of the barotropic streamfunction $\psi$ (centered around their means). Solid lines indicate the real part and dashed lines the imaginary parts. . . . .	29
Figure 2.9	Damping time, $1/d$ of the MSM. . . . .	29
Figure 2.10	The true barotropic streamfunction (left) compared with the filtered linear interpolation scheme (middle) and the filtered kriging scheme (right), with $M = 36$ and $r^o = 6.9$ at time $T = 340$ . The circles depict observation locations. . . . .	32
Figure 2.11	The true barotropic streamfunction (left) compared with the filtered linear interpolation scheme (middle) and the filtered kriging scheme (right), with $M = 18$ and $r^o = 6.9$ at time $T = 340$ . The circles depict observation locations. . . . .	33
Figure 2.12	The true barotropic streamfunction (left) compared with the filtered linear interpolation scheme (middle) and the filtered kriging scheme (right), with $M = 49$ and $r^o = 6.9$ at time $T = 340$ . The circles depict observation locations. . . . .	34
Figure 2.13	The true barotropic streamfunction (left) compared with the filtered linear interpolation scheme (middle) and the filtered kriging scheme (right), with $M = 36$ and $r^o = 17.3$ at time $T = 340$ . The circles depict observation locations. . . . .	34
Figure 2.14	The RMS errors (left panels) and correlations (right panels) for each case. Unfiltered kriging (dashes), filtered kriging (dashes with ‘+’ sign), unfiltered linear interpolation (solid line), and filtered linear interpolation (solid line with circles). . . . .	35



Figure 2.15	The average RMS errors (left panels) and pattern correlations (right panels) for numerical experiments with regularly spaced observations at the $N \times N = 36$ model grid points, with noise variance $r^o = 10\%E$ (top panels) and $r^o = 25\%E$ (bottom panels). . . . .	36
Figure 2.16	The RMS errors as functions of time, with $M = N^2 = 36$ , $r^o = 17.3$ and $T_{obs} = 0.01$ . The top panel shows the RMS error of the unfiltered linearly interpolated estimates and the bottom panel shows the RMS error of unfiltered kriging estimates. . . . .	37
Figure 2.17	The truth (top panel) interpolated (middle panels) and filtered (bottom panels) fields at time $T = 363.28$ (in contours). The circles depict observation locations. . . . .	38
Figure 2.18	The truth, (top panel), interpolated (middle panels) and filtered (bottom panels) fields at time $T = 61$ (in contours). The circles depict observation locations. . . . .	39
Figure 3.1	Contour plot of deep convective heating $P$ from a numerical simulation of the multicloud model. . . . .	51
Figure 3.2	Weighting functions for three AIRS channels. . . . .	55
Figure 3.3	The switch function $\Lambda$ (top panel) and the normalized deep convective, stratiform, and congestus heating rates, which correspond to the cloud fraction (bottom panel), at location $x = 20,000$ over time. . . . .	64
Figure 3.4	Clear (left) and cloudy (right) AIRS observations at one location, over time. The top panels correspond to channel frequency $680.142 \text{ cm}^{-1}$ , the middle panels correspond to channel frequency $1377.426 \text{ cm}^{-1}$ and the bottom panels correspond to channel frequency $2500.602 \text{ cm}^{-1}$ . . . . .	65
Figure 3.5	Congestus cloud (top), stratiform cloud (middle) and deep convective cloud (bottom) effects on brightness temperatures for all channels. The black lines represent the clear sky pCRTM radiances while gray lines represent the cloudy sky radiances. . . . .	66
Figure 3.6	Data assimilation of satellite observations with the multicloud model. . . . .	67
Figure 3.7	Weighting functions of AIRS channels that are minimally obstructed by congestus clouds. The gray area represents the atmospheric region below a congestus cloud. . . . .	72
Figure 3.8	Weighting functions of AIRS channels that are minimally obstructed by deep convective clouds and stratiform clouds. The gray area shows the atmospheric region below their tops. . . . .	73
Figure 3.9	Weighting functions of AIRS channels that observe the entire atmosphere. . . . .	74
Figure 3.10	Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at time $T = 1,000$ days, with direct observations of temperature and humidity, $D_{obs} = 40 \text{ km}$ , $T_{obs} = 24 \text{ hours}$ , $K = 16$ , covariance inflation $r = 19\%$ and localization radius $80 \text{ km}$ . . . . .	75

Figure 3.11	Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at location $x = 20,000$ km over the last 30 days, with direct observations of temperature and humidity, $D_{obs} = 40$ km, $T_{obs} = 24$ hours, $K = 16$ , covariance inflation $r = 19\%$ and localization radius 80 km. . . . .	76
Figure 3.12	Spatially and temporally averaged RMS error and temporally averaged pattern correlation between the posterior and prior state estimates of $q$ and $P$ and true signal as a function of localization radius, with direct observations of the temperature and humidity, $K = 16$ , and $T_{obs} = 24$ hours. The black lines correspond to posterior state estimates, while the gray lines correspond to prior state estimates. . . . .	77
Figure 3.13	RMS errors between the posterior state estimates of $q$ and $H_c$ and their true signals as a function of ensemble size, with clear sky observations, $D_{obs} = 80$ km, $T_{obs} = 24$ hours, and localization radius 80 km. . . . .	78
Figure 3.14	Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at time $T = 1,000$ days, with clear sky observations, $D_{obs} = 80$ km, $T_{obs} = 24$ hours, $K = 16$ , covariance inflation $r = 16\%$ and localization radius 80 km. . . . .	79
Figure 3.15	Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at location $x = 20,000$ km over the last 30 days, with clear sky observations, $D_{obs} = 80$ km, $T_{obs} = 24$ hours, $K = 16$ , covariance inflation $r = 16\%$ and localization radius 80 km. . . . .	80
Figure 3.16	Spatially and temporally averaged RMS error and temporally averaged pattern correlation between the posterior (black lines) and prior (gray lines) state estimates of $q$ and $P$ and true signal, as a function of localization radius, with $K = 16$ , $T_{obs} = 24$ and clear sky AIRS observations. . . . .	81
Figure 3.17	Spatially averaged RMSE between the posterior state estimates of $u_1$ , $\theta_2$ , $q$ and $P$ and their true signals over time with $K = 16$ , $T_{obs} = 24$ hours, and localization radius 80 km. . . . .	82
Figure 3.18	Spatially averaged RMSE between the posterior state estimates of $u_1$ , $\theta_2$ , $q$ and $P$ and their true signals over time with $K = 16$ , $T_{obs} = 24$ hours, and localization radius 80 km. . . . .	84
Figure 3.19	Spatially averaged RMSE between posterior state estimates of $u_1$ , $\theta_2$ , $q$ and $P$ and their true signals over time with $K = 16$ , $T_{obs} = 24$ hours, and localization radius 80 km. . . . .	85
Figure 3.20	Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at time $T = 1,000$ days, with cloudy sky AIRS observations, $D_{obs} = 40$ km, $T_{obs} = 24$ hours, $K = 16$ , covariance inflation $r = 14\%$ and localization radius 80 km. . . . .	86
Figure 3.21	Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at location $x = 20,000$ km over the last 30 days, with cloudy sky AIRS observations, $D_{obs} = 40$ km, $T_{obs} = 24$ hours, $K = 16$ , covariance inflation $r = 14\%$ and localization radius 80 km. . . . .	87

Figure 3.22	Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at location $x = 20,000$ km over the last 30 days, with cloudy sky AIRS observations and direct observations of $H_s$ , $D_{obs} = 40$ km, $T_{obs} = 24$ hours, $K = 16$ and localization radius 80 km. . . . .	88
Figure 3.23	The true precipitation contours (top), and estimated precipitation contours (bottom). The bottom right panel shows the result of using the true value of $H_s$ in the filter. Here, $T_{obs} = 6$ hours, $D_{obs} = 40$ km and $\mathcal{D} = 2$ . . . . .	89
Figure 3.24	Spatially and temporally averaged RMS error between the posterior (black lines) and prior (gray lines) state estimates of the multicloud model prognostic variables and true signal as a function of $D_{obs}$ , with $K = 16$ , $T_{obs} = 24$ hours, and localization radius 80 km. . . . .	90
Figure 3.25	Spatially and temporally averaged RMS error between the posterior (black lines) and prior (gray lines) state estimates and true signal as a function of $D_{obs}$ , with $K = 16$ , $T_{obs} = 24$ hours, and localization radius 80 km. . . . .	91
Figure 3.26	Temporally and Spatially averaged RMS error and temporally averaged pattern correlation between the posterior (black lines) and prior (gray lines) state estimates and true signal as a function of localization radius, with cloudy sky AIRS observations, $T_{obs} = 24$ hours, and $K = 16$ . . . . .	92

# Chapter 1

## Introduction

Filtering is the process of determining the statistically best estimate of a true signal based on partial, noisy observations. It is a two-step predictor-corrector scheme consisting of a forecast step, followed by an analysis step that updates the prediction as observations become available. The forecast treats the underlying system dynamics as a black box, while the analysis accounts for both the prediction and the observations to update the state estimate [45]. Consider a random variable  $\vec{u}_m$  whose dynamics are governed by the discrete time model

$$\vec{u}_{m+1} = \vec{f}(\vec{u}_m) + \vec{\sigma}_{m+1}, \quad (1.1)$$

where  $\vec{f}$  is a discrete time operator that propagates  $\vec{u}_m \in \mathbb{C}^N$  or  $\mathbb{R}^N$  forward in time and  $\vec{\sigma}_m \in \mathbb{C}^N$  or  $\mathbb{R}^N$  is a Gaussian noise vector of mean zero and covariance

$$\langle \vec{\sigma}_m \otimes (\vec{\sigma}_m)^* \rangle = R.$$

The superscript asterisk denotes the conjugate transpose, and the notation  $\langle \bullet \rangle$  denotes expectation with respect to the Gaussian noise joint distribution.

The goal of the Kalman filter is to estimate the true state  $\vec{u}_{m+1}$ , given noisy observations  $\vec{v}_{m+1}$  that satisfy the relation

$$\vec{v}_{m+1} = \vec{g}(\vec{u}_{m+1}) + \vec{\sigma}_{m+1}^o, \quad (1.2)$$

where  $\vec{g} \in \mathbb{R}^{M \times N}$  is an observation operator that maps the model variable  $\vec{u}_{m+1}$  into the observational space and  $\vec{\sigma}_{m+1}^o \in \mathbb{C}^M$  or  $\mathbb{R}^M$  is a Gaussian measurement error of mean zero and diagonal covariance

$$\langle \vec{\sigma}_m^o \otimes (\vec{\sigma}_m^o)^* \rangle = r^o \mathcal{I}.$$

The derivation of the Kalman filter assumes Gaussian noise. For certain applications, this assumption may not be true; in this case, one might consider another filtering method, such as

a particle filter.

We will examine two different filtering approaches in two settings. In the first setting, we consider observations of two-dimensional turbulent ocean dynamics in the mid-latitude. Let  $u_{k,l,m}$  denote the  $(k, l)$ -Fourier mode of  $\vec{u}_m$ . This filter is implemented mode-by-mode. Therefore, to simplify the discussion, we will drop the  $k$  and  $l$  in the subscript of  $u_{k,l,m}$ . Here we model the underlying dynamics in a Fourier domain with a simple model

$$u_{m+1} = Fu_m + \sigma_{m+1}. \quad (1.3)$$

Using the two-layer quasi-geostrophic (QG) model, [61, 62], we generate numerous two dimensional sparse, noisy, irregularly spaced observation networks. This imitates the practical situation in which physical observations are sparse and not available at regular model grid points. We simulate the QG model, using parameters that mimic oceanic turbulence and we only take our observations from its large scale components, making this a difficult filtering problem. Our goal is to assess the filtering skill when an irregularly spaced observation network is subject to processing prior to assimilation. The filtering approach here consists of a two step hierarchical Bayesian framework. The first step is to interpolate the irregularly spaced observations to the regular filter model grid points. The second step is to apply a simple, reduced filter stochastic filter in Fourier space. In this set up, the model in (1.2) simplifies to,

$$v_{m+1} = Gu_{m+1} + \sigma_{m+1}^o,$$

where  $v_{m+1}$  is the Fourier mode of  $\vec{v}_{m+1}$  and  $\sigma_{m+1}^o$  denotes the observational noise with variance  $(r^o/M)\mathcal{I}$  (see [45]). The component of  $G$  corresponding to mode  $(k, l)$  is equal to 1, while the other components of  $G$  are equal to 0. We will use a numerically cheap forecast model, the Mean Stochastic Model, [45, 48] to parametrize our forecast operator  $F$  in (1.3).

In this application, we use a Kalman filter, assuming that the observation and prior forecast error uncertainties are Gaussian, uncorrelated, and unbiased [45]. The error distribution of the observations  $v_{m+1}$  is given by the Gaussian conditional distribution

$$P(v_{m+1}|u_{m+1}) \sim \mathcal{N} \left( Gu_{m+1} \left| \frac{r^o}{M} \mathcal{I} \right. \right).$$

We denote the estimate of  $u_{m+1}$  before updating to account for the observations  $v_{m+1}$  at this time step as  $u_{m+1}^b$ . This estimate is called the prior state estimate, forecast, or background estimate, and we use the three terms interchangeably. Then

$$u_{m+1}^b = Fu_m^a + \sigma_{m+1},$$

where  $u_m^a$  denotes the posterior state estimate, or analysis at the previous time step. The posterior state  $u_{m+1}^a$  gives the estimate of the true state after considering the observations  $v_{m+1}$ .

Probabilistically, the prior state probability distribution of  $u_{m+1}^b$  accounts for all observations up to time  $t_m$  as

$$P(u_{m+1}^b) \sim \mathcal{N}(\bar{u}_{m+1}^b, R_{m+1}^b).$$

Here  $\bar{u}_{m+1}^b \equiv \langle u_{m+1}^b \rangle$  gives the mean of the prior state and

$$R_{m+1}^b \equiv \langle (u_{m+1}^b - \bar{u}_{m+1}^b)(u_{m+1}^b - \bar{u}_{m+1}^b)^* \rangle$$

gives the prior error covariance. The estimate for the posterior mean is then found by maximizing the conditional distribution

$$P(u_{m+1}^a) \equiv P(u_{m+1}^b | v_{m+1}) \propto P(u_{m+1}^b) P(v_{m+1} | u_{m+1}^b).$$

We then obtain the posterior error covariance by computing

$$R_{m+1}^a = \langle (u_{m+1} - \bar{u}_{m+1}^a)(u_{m+1} - \bar{u}_{m+1}^a)^* \rangle.$$

In the second part of this thesis, we assimilate synthetic cloudy Atmospheric Infrared Sounder (AIRS) satellite observations equally spaced along the equator with a toy model of tropical convection called the multicloud model [32, 33]. As we will demonstrate, the presence of a cloud can have a significant effect on satellite observations. Less than 1% of AIRS observations are taken over a cloud-free atmospheric column [13]. Consequently, only a small percentage of AIRS observations are currently used in forecast models. We will study the potential and the merit of assimilating cloud-impacted AIRS data, and examine the biases that clouds can introduce in the filtering algorithm. We use a simulation of the multicloud model to create a true atmospheric state. To generate synthetic satellite observations, we use the prototype community Radiative Transfer Model, [22], augmented with added cloud effects. In this case, the forecast model  $\vec{f}$  in (1.1) represents the multicloud and is a nonlinear, deterministic operator. The observation operator  $\vec{g}$  in (1.2) is also nonlinear and is a composite of the radiative transfer model, an inverse Planck function, and several dimensionalization transformations. Notice that because we use the multicloud model to specify the true dynamics, the filter model is perfectly specified. Therefore, this is an example of a perfect model experiment [9].

The basic Kalman filter assumes linearity in both the forecast model in (1.1) and the observation operator in (1.2) in its derivation. Because of the nonlinearity of both the multicloud model and the radiative transfer model, we must be careful to employ an appropriate filtering

strategy. Here we utilize the local ensemble transform Kalman filter (LETKF), a deterministic approach, in the analysis step [28, 7]. This filter is implemented locally, meaning that for each model grid point, we use all of the observations available within a radius  $\mathcal{D}$  in the analysis update. The LETKF uses a collection, or ensemble, of state vectors to represent the distribution of the true state,  $u_{m+1}$ . Ensemble members are created with different initial conditions. Starting with an ensemble  $\{\vec{u}_k^a\}_{k=1}^K$  at time  $t_m$ , the filter propagates each ensemble member forward in time with the nonlinear forecast model  $\vec{f}$  to obtain  $\{\vec{u}_k^b\}_{k=1}^K$  at time  $t_{m+1}$ .

When observations become available, the analysis step uses the forecast ensemble  $\{\vec{u}_k^b\}$ , the local observations and their error covariance matrix, and the observation operator  $\vec{g}$  to compute the posterior ensemble mean and posterior error covariance matrix. The filter assumes that the forecast (prior) ensemble mean  $\bar{\vec{u}}^b$  gives the best estimate of the true atmospheric state, before taking new observations into account. In the analysis, each ensemble member is updated with the analysis (posterior) ensemble mean,  $\bar{\vec{u}}^a$ , which represents the best estimate of the atmospheric state after considering the observations.

The forecast error covariance depends on the perturbation of each ensemble member about the prior ensemble mean,  $\vec{u}_k^b - \bar{\vec{u}}^b$ . The basic idea of the LETKF is to apply a transformation to the prior ensemble perturbations so that they describe the posterior error covariance that accounts for the observations. In a sense, the perturbations are weighted, based on the difference between the average of  $\{\vec{g}(\vec{u}_k^b)\}$  and the values of the local observations. The transformed perturbations are used to define the posterior error covariance, and to update each ensemble member in conjunction with the posterior mean.

The remainder of this thesis is outlined as follows. In Chapter 2, we consider a hierarchical Bayesian filtering scheme. This study has been published in [8]. We first discuss the two-layer QG model in the mid-latitude. Next, we discuss the two-step filtering approach; in the first step, we consider two interpolation techniques, a two-dimensional piecewise linear interpolation scheme and ordinary kriging, [10], a geostatistical interpolation scheme. In the second step, we use the Mean Stochastic Model [45]. Subsequently, we present the results of several filtering experiments and close the chapter with a summary and discussion. In Chapter 3, we assimilate cloudy satellite observations. We first discuss the multicloud model and the radiative transfer model. Next, we explain how we create satellite observations, and how we simulate cloud cover. We then discuss the filtering approach. Finally, we present our numerical results and close with a summary and discussion.

## Chapter 2

# A Hierarchical Bayesian Reduced Stochastic Filter

In this chapter, we consider a filtering approach for assimilating noisy, irregularly spaced, sparsely observed turbulent signals with a hierarchical Bayesian reduced stochastic filter. Typical forecast models are resolved on regularly spaced grid points. In contrast, observations in nature are usually sparse and noisy, and are not always available at regularly spaced grid points. For example, radiosonde measurements from weather balloons are sparse over the ocean, but denser over land. Our goal here is to extend the current reduced stochastic filtering models to account for sparse, irregularly spaced observation networks on a two-dimensional domain, through the use of an appropriate interpolation technique.

Let  $\tilde{V}$  denote the random variable of the irregularly spaced observations, and let  $U$  denote the random variable of the forecast model state. The standard stochastic filtering approach utilizes Bayes' theorem to obtain the posterior statistics of the conditional distribution

$$P(u|\tilde{v}) \propto P(u)P(\tilde{v}|u), \quad (2.1)$$

where  $\tilde{v} \in \tilde{V}$  and  $u \in U$  [2, 7, 14]. Here,  $P(u)$  gives the prior distribution associated with the forecast model dynamics. The typical method for assimilating irregularly spaced observations is to use the observation operator

$$\tilde{v}_m = G(u_m) + \tilde{\sigma}_m^o, \quad \tilde{\sigma}_m^o \sim \mathcal{N}(0, r^o \mathcal{I}) \quad (2.2)$$

to map the model variable  $u_m = (u_{j,m})_{j=1}^N$  at grid point  $j$  and discrete time step  $m$  to the observation variable  $\tilde{v}_m = (\tilde{v}_{\ell,m})_{\ell=1}^M$  at grid point  $\ell$ . Here, the observation  $\tilde{v}_m$  is corrupted by an i.i.d. Gaussian noise  $\tilde{\sigma}_m^o$  with mean zero and variance  $r^o$ . In (2.1),  $P(\tilde{v}|u)$  gives a likelihood function associated with this observation model..



The hierarchical approach can be probabilistically interpreted as an application of the hierarchical Bayesian approach [23]. The first step in this approach is to use a statistical interpolation technique to interpolate the irregularly spaced observations  $\tilde{v} \in \tilde{V}$  to the filter model grid points. Let  $V$  denote the random variable of the interpolated observations. Then from Bayes' theorem, we obtain

$$P(u|v) \propto P(u)P(v|u),$$

where  $v \in V$ . Adding the condition of  $v$  to (2.1), we have

$$P(u|\tilde{v}, v) \propto P(u|v)P(v|u, \tilde{v}) \propto P(u)P(v|u)P(\tilde{v}|v, u). \quad (2.3)$$

The conditional distribution  $P(v|u, \tilde{v})$  represents the outcome of the statistical interpolation. The second step of this approach is to apply  $P(u)P(v|u, \tilde{v})$  through a reduced stochastic filtering algorithm. Here, the computational cost is significantly reduced compared to the standard Bayesian approach in (2.1) with the observation model in (2.2). This process combines the data-driven interpolation in the first step and the dynamical constraint of the prior distribution in the second step.

We will apply the hierarchical Bayesian approach in (2.3) to filter noisy, irregularly spaced, sparse observations of geophysical turbulence over a two-dimensional doubly periodic domain. The two layer quasi-geostrophic (QG) model, [61, 62], simulates atmospheric and oceanic turbulence. We generate our observations from the large-scale, barotropic components of a solution of the QG model, using a long Rossby radius, corresponding with ocean dynamics. In reality, observations of small scale processes are typically not available. Additionally, we externally force the QG model, so that it exhibits baroclinic instability. We choose the locations of the observations randomly, and perturb each with a noise  $r^o$ .

In the first step of (2.3), we will compare the performance of two different interpolation schemes on numerous observation networks. In particular, we compare a deterministic piecewise linear interpolation and ordinary kriging, a spatial statistical interpolation [10]. In the linear interpolation, we use three observations to interpolate at each model grid point. This choice is arbitrary; compared to a one-dimensional Lagrange function, which can only be defined with two observations, a two dimensional Lagrange function can be defined with any number of surrounding observations. Ordinary kriging is a data driven interpolation technique. This method treats the set of observations and interpolated observations as a random field, and assumes Gaussian, stationary noises. The main idea of kriging is to fit the set of observations to an empirically chosen, parametric covariance function. The expected value at each model grid point is then found by solving a constrained minimization problem.

The assimilation of sparsely observed turbulent signals has been studied previously in a one-dimensional setting in [23]. As in [23], we will use the Mean Stochastic Model (MSM),

[45, 48], a filtering strategy that models turbulent fluctuations with a linear damping and white noise, [61, 50, 51, 43, 11, 47], for the second step in (2.3). This cheap filter is comprised of block diagonal Langevin equations in Fourier space. We obtain the MSM parameters offline by fitting a training data set to the solutions of the Ornstein-Uhlenbeck process. For problems similar to ours, the MSM has produced positive results, despite its simplicity [25, 26, 37].

In Section 2.1, we will review the two-layer quasi-geostrophic model, [61, 62], to be used in our numerical experiments. In Section 2.2, we will discuss the interpolation schemes and the steps in the hierarchical Bayesian approach in detail. We report the filtering results in Section 2.3 and summarize our findings in Section 2.4.

## 2.1 Quasi-geostrophic Model

We consider observations of turbulent ocean dynamics in the mid-latitude, between  $30^\circ$  and  $65^\circ$  in the northern hemisphere. To generate our observations, we use a numerical solution of the two-layer quasi-geostrophic (QG) model in a two-dimensional periodic domain, with instability induced by mean vertical shear [61].

### 2.1.1 Background Information

In the mid-latitude, the Coriolis force, a pseudo-force associated with the effects of Earth's rotation, affects geophysical fluid dynamics. This force is characterized by the Coriolis parameter  $f$ , which is proportional to Earth's angular velocity and increases with increasing latitude (note that  $f = 0$  at the equator). Using a tangent plane approximation, we can approximate the Coriolis parameter by

$$f = f_0 + \beta y,$$

where the  $y$  direction points towards the north pole for the northern hemisphere, and  $f_0$  depends on the reference latitude.

The QG model is derived from the shallow water equations under several assumptions. In the presence of Coriolis force  $\vec{f} = f\mathbf{k}$ , the single layer shallow water equations are given in [65] by

$$\begin{aligned} \frac{\partial \vec{u}}{\partial t} + \vec{u} \cdot \nabla \vec{u} + \vec{f} \times \vec{u} &= -g \nabla_z \eta \\ \frac{Dh}{Dt} + \nabla \cdot (h\vec{u}) &= 0, \end{aligned} \tag{2.4}$$

where  $\vec{u}(x, y, t)$  is the horizontal velocity,  $\eta(x, y, t)$  is the free surface height,  $h(x, y, t)$  is the

fluid thickness,  $g$  is the gravity constant, and

$$\nabla_z = \frac{\partial}{\partial x} \mathbf{i} + \frac{\partial}{\partial y} \mathbf{j}.$$

The material derivative  $D/Dt$  of a function  $\phi$  that is advected by  $\vec{u}$  is defined by

$$\frac{D\phi}{Dt} \equiv \frac{\partial\phi}{\partial t} + \vec{u} \cdot \nabla\phi.$$

For a flat-bottom system,  $h(x, y, t) = \eta(x, y, t)$ . The first equation is (2.4) describes the momentum of the fluid, while the second equation represents mass conservation.

We can quantify the significance of rotation in a fluid with the Rossby number, which is given by

$$R_o \equiv \frac{U}{fL},$$

where  $L$  is the horizontal length scale and  $U$  is the horizontal velocity scale. If the Rossby number is small, the effects of rotation are important, as is the case in large-scale oceanic flow. In this case, the Coriolis force dominates the advection terms in (2.4). The only term in which can balance the Coriolis term in (2.4) is the free surface gradient. Therefore, we define geostrophic balance as

$$\vec{f} \times \vec{u} \approx -g\nabla_z\eta.$$

In the QG equations, the deformation radius  $L_d$ , the scale at which Earth's rotation becomes significant to fluid dynamics, is the horizontal length scale. Thus, the QG approximation assumes a small Rossby number and hence that the fluid is in geostrophic balance. Additionally in this setting, we assume that variations in the Coriolis parameter are small, therefore  $|\beta L| \ll |f_o|$ . Finally, we assume that time scales advectively, so that we can use the time scale  $T = L_d/U$ .

### 2.1.2 The Model

Consider a two-layer shallow water system with a flat bottom and rigid lid, and denote fluid thickness in each layer as  $h_i(x, y, t)$  and the velocity in each layer as  $\vec{u}_i = (u_i, v_i, 0)$ . The vorticity  $\omega_i$  in each layer is defined to be the curl of the velocity field,  $\omega_i = \nabla \times \vec{u}_i$ . Because of the shallow water approximation, only the vertical component of vorticity is nonzero, and therefore the vorticity simplifies to

$$\omega_i = \left( \frac{\partial v_i}{\partial x} - \frac{\partial u_i}{\partial y} \right) = \varsigma_i \mathbf{k}.$$

In a rotating shallow water system, the quantity

$$Q \equiv \frac{\varsigma + f}{h}$$

is called the potential vorticity. Assume that both layers are of equal mean thickness  $H$  and let  $h_i = H + h'_i$ . Because the variations in the layer thickness are small, we can Taylor expand  $Q_i$  in each layer about  $h'_i$ . Taking a first order approximation yields

$$\begin{aligned} Q_i &= \left( \frac{\varsigma_i + f}{H} \right) \left( \frac{1}{1 + \frac{h'_i}{H}} \right) \\ &\approx \left( \frac{\varsigma_i + f}{H} \right) \left( 1 - \frac{h'_i}{H} \right) \\ &\approx \frac{1}{H} \left( \varsigma_i + f - \varsigma_i \frac{h'_i}{H} - f \frac{h'_i}{H} \right) \end{aligned}$$

Because the system is in geostrophic balance, the Coriolis terms and  $\varsigma_i$  dominate this equation. Therefore,

$$Q_i \approx \frac{1}{H} \left( \varsigma_i + f - f \frac{h'_i}{H} \right).$$

The variations in  $f$  are small, as are the variations in layer thickness, so we can further approximate

$$Q_i \approx \frac{1}{H} \left( \varsigma_i + f - f_o \frac{h'_i}{H} \right).$$

In the QG equations,  $Q_i$  is advected, therefore only non-constant terms are important to the model dynamics. Since  $f_o$  and  $H$  are constant, we therefore define the quasi-geostrophic potential vorticity in each layer by

$$q_i = \beta y + \varsigma_i - f_o \frac{h'_i}{H}.$$

In each layer, we define a streamfunction  $\psi_i$  by

$$\psi_1 = \frac{g}{f_o}(h'_1 + h'_2), \quad \psi_2 = \frac{g}{f_o}(h'_1 + h'_2) + \frac{g'h'_2}{f_o},$$

where  $g' = (\rho_2 - \rho_1)/\rho_1$ , subscript 1 denotes the upper layer and subscript 2 denotes the bottom layer [65]. These are defined so that

$$u_i = -\frac{\partial \psi_i}{\partial y}, \quad v_i = \frac{\partial \psi_i}{\partial x}, \quad \varsigma_i = \nabla^2 \psi_i.$$

Thus the QG potential vorticity in each layer becomes

$$\begin{aligned} q_1 &= \nabla^2 \psi_1 + \frac{f_o^2}{g'H}(\psi_2 - \psi_1) + \beta y + \frac{f_o^2}{gH_2} \psi_1 \\ q_2 &= \nabla^2 \psi_2 + \frac{f_o^2}{g'H}(\psi_1 - \psi_2) + \beta y. \end{aligned}$$

The last term in the first equation is neglected in the rigid lid approximation, because  $g \gg g'$ . Nondimensionalizing these equations, using the deformation radius  $L_d$  as the horizontal scale of motion, we obtain

$$q_i = \nabla^2 \psi_i + \frac{k_d^2}{2}(\psi_{3-i} - \psi_i) + \beta y, \quad j = 1, 2 \quad (2.5)$$

where  $k_d = \sqrt{8}/L_d$  gives the wavenumber associated with  $L_d$ .

To derive the QG equations in each layer, we start with the nondimensionalized momentum equation in geostrophic balance in layer  $i$ ,

$$\vec{f} \times \vec{u}_i = -\nabla_z \eta_i. \quad (2.6)$$

Here,  $L$  is the horizontal length scale,  $U$  is the horizontal velocity scale, and  $T = U/L$  is the advective time scale. To nondimensionalize the Coriolis force, we divide by  $f_o$ . Then the dimensionless Coriolis parameter is

$$f = 1 + R_o \hat{\beta} y,$$

where  $\beta = \hat{\beta}U/L^2$ . The momentum equation then becomes

$$\begin{aligned} \vec{f} \times \vec{u}_i &= -\nabla_z \eta_i \\ -v_i(1 + R_o \hat{\beta} y)\mathbf{i} + u_i(1 + R_o \hat{\beta} y)\mathbf{j} &= -\frac{\partial \eta_i}{\partial x}\mathbf{i} + \frac{\partial \eta_i}{\partial y}\mathbf{j}. \end{aligned}$$

Following [65], we expand the variables  $u$ ,  $v$  and  $\eta$  in an asymptotic series, using  $R_o$  as the small parameter, and substitute them into the momentum equation. Then

$$u_i = u_i^0 + R_o u_i^1 + R_o^2 u_i^2 + \dots \quad v_i = v_i^0 + R_o v_i^1 + R_o^2 v_i^2 + \dots$$

and

$$\eta_i = \eta_i^0 + R_o \eta_i^1 + \dots$$

We first substitute the expansions for  $u$  and  $v$  into (2.6) and equate the powers of  $R_o$ . At the lowest order, this yields the equalities

$$u_i^0 = -\frac{\partial \eta_i^0}{\partial y} \quad v_i^0 = \frac{\partial \eta_i^0}{\partial x}.$$

Cross-differentiating, we obtain

$$\nabla \cdot \vec{u}_i^0 = 0.$$

The next order momentum equation is given by

$$\frac{\partial \vec{u}_i^0}{\partial t} + (\vec{u}_i^0 \cdot \nabla) \vec{u}_i^0 + \hat{\beta} y \mathbf{k} \times \vec{u}_i^0 - \mathbf{k} \times \vec{u}_i^1 = -\nabla \eta_i^1.$$

Taking the curl of this, we obtain the vorticity

$$\frac{\partial \varsigma_i^0}{\partial t} + (\vec{u}_i^0 \cdot \nabla)(\varsigma_i^0 + \hat{\beta} y) = -\nabla \cdot \vec{u}_i^0, \quad (2.7)$$

where

$$\varsigma_i^0 = \frac{\partial v_i^0}{\partial x} - \frac{\partial u_i^0}{\partial y}.$$

The nondimensionalized mass continuity equation is given by

$$R_o F \frac{D\eta_i}{Dt} + (1 + R_o F \eta_i) \nabla \cdot \mathbf{u}_i = 0, \quad (2.8)$$

with  $\sqrt{F} = L/L_d$ . At the lowest order, the mass equation simplifies to  $\nabla \cdot \vec{u}_i^0 = 0$ , as the lowest order momentum equation did. At the next order, we have

$$F \frac{\partial \eta_i^0}{\partial t} + F \vec{u}_i^0 \cdot \nabla \eta_i^0 + \nabla \cdot \vec{u}_i^1 = 0. \quad (2.9)$$

Substituting (2.9) into (2.7) gives

$$\frac{\partial \varsigma_i^0}{\partial t} + (\vec{u}_i^0 \cdot \nabla)(\varsigma_i^0 + \hat{\beta} y) = F \frac{\partial \eta_i^0}{\partial t} + F \vec{u}_i^0 \cdot \nabla \eta_i^0,$$

or

$$\frac{\partial}{\partial t}(\varsigma_i^0 - F \eta_i^0) + (\vec{u}_i^0 \cdot \nabla)(\varsigma_i^0 + \hat{\beta} y - F \eta_i^0) = 0.$$

With the nondimensionalized streamfunction  $\psi_i$ , this becomes

$$\frac{\partial}{\partial t}(\nabla^2 \psi_i^0 - F \psi_i^0) + (\vec{u}_i^0 \cdot \nabla)(\varsigma_i^0 + \hat{\beta} y - F \psi_i^0) = 0,$$

or

$$\frac{D^0}{Dt}(\nabla^2 \psi_i^0 + \hat{\beta} y - F \psi_i^0) = 0,$$

where the superscript 0 in the material derivative indicates that the lowest order velocity,  $\vec{u}_i^0$  is

the advecting velocity. Restoring the dimensions, this yields

$$\frac{D}{Dt}(\nabla^2\psi_i + \beta y - \frac{1}{L_d}\psi_i) = \frac{Dq_i}{Dt} = 0.$$

Expanding the material derivative, and using the properties  $u = -\psi_y$  and  $v = \psi_x$ , we obtain the QG equation in each layer:

$$\begin{aligned} \frac{Dq_i}{Dt} &= \frac{\partial q_i}{\partial t} + \vec{u}_i \cdot \nabla q_i &= \frac{\partial q_i}{\partial t} + u_i \frac{\partial q_i}{\partial x} + v_i \frac{\partial q_i}{\partial y} \\ &= \frac{\partial q_i}{\partial t} - \frac{\partial \psi_i}{\partial y} \frac{\partial q_i}{\partial x} + \frac{\partial \psi_i}{\partial x} \frac{\partial q_i}{\partial y} \\ &= \frac{\partial q_i}{\partial t} + J(\psi_i, q_i). \end{aligned}$$

The Jacobian function  $J(\psi, q) = \psi_x q_y - \psi_y q_x$  in

$$\frac{\partial q_i}{\partial t} + J(\psi_i, q_i) = 0 \quad (2.10)$$

represents nonlinear advection.

Because we consider a two-layer QG model that is forced by a mean vertical shear, we take  $\psi_1$  and  $\psi_2$  to be perturbed streamfunctions with background state

$$\Psi_1 = -Uy, \quad \Psi_2 = Uy. \quad (2.11)$$

Here  $U$  is the zonal mean shear, selected so that this model exhibits baroclinic instability with a turbulent cascade. For details on the properties of this system, see [61] and the citations in [62]. Substituting  $\psi_1 = \Psi_1 + \psi'_1$ , and  $\psi_2 = \Psi_2 + \psi'_2$  into (2.5) yields

$$\begin{aligned} q'_1 &= \nabla^2 \psi'_1 + \frac{k_d^2}{2}(\psi'_2 - \psi'_1) + (k_d^2 U + \beta)y \\ q'_2 &= \nabla^2 \psi'_2 + \frac{k_d^2}{2}(\psi'_1 - \psi'_2) + (\beta - k_d^2 U)y. \end{aligned}$$

We substitute both of these into (2.10). The equation for  $q'_1$  simplifies as

$$\begin{aligned} \frac{\partial q'_1}{\partial t} + J(\psi_1, q'_1) &= \frac{\partial q_1}{\partial t} + \left( \frac{\partial \psi'_1}{\partial x} \right) \left( \frac{\partial q_1}{\partial y} + k_d^2 U + \beta \right) - \left( \frac{\partial \psi'_1}{\partial y} - U \right) \left( \frac{\partial q_1}{\partial x} \right) \\ &= \frac{\partial q_1}{\partial t} + J(\psi'_1, q_1) + (k_d^2 U + \beta) \frac{\partial \psi'_1}{\partial x} + U \frac{\partial q_1}{\partial x}, \end{aligned}$$

while the equation for  $q'_2$  simplifies as

$$\begin{aligned}\frac{\partial q'_2}{\partial t} + J(\psi_2, q'_2) &= \frac{\partial q_2}{\partial t} + \left(\frac{\partial \psi'_2}{\partial x}\right) \left(\frac{\partial q_2}{\partial y} - k_d^2 U + \beta\right) - \left(\frac{\partial \psi'_2}{\partial y} + U\right) \left(\frac{\partial q_2}{\partial x}\right) \\ &= \frac{\partial q_2}{\partial t} + J(\psi'_2, q_2) + (\beta - k_d^2 U) \frac{\partial \psi'_2}{\partial x} - U \frac{\partial q_2}{\partial x}.\end{aligned}$$

The lower layer experiences a friction at its bottom; the effects of Earth's rotation on large scale oceanic flow is large, and results in a boundary layer, called the Ekman layer, which balances the Coriolis force with friction [65]. In the lower level vorticity equation, Ekman friction is represented by the additional term  $\kappa \nabla^2 \psi'_2$ , where  $\kappa$  is the Ekman bottom drag coefficient.

Dropping the primes on  $\psi'_i$  to simplify the notation, two-layer QG equations with a flat bottom, and rigid lid are given in [8] by

$$\begin{aligned}\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x}\right) q_1 + J(\psi_1, q_1) + \frac{\partial \psi_1}{\partial x} (\beta + k_d^2 U) + \nu \nabla^8 q_1 &= 0, \\ \left(\frac{\partial}{\partial t} - U \frac{\partial}{\partial x}\right) q_2 + J(\psi_2, q_2) + \frac{\partial \psi_2}{\partial x} (\beta - k_d^2 U) + \kappa \nabla^2 \psi_2 + \nu \nabla^8 q_2 &= 0.\end{aligned}\quad (2.12)$$

Notice here the additional  $\nu \nabla^8 q_i$  term in each equation; here  $\nu$  is a hyperviscosity coefficient, chosen so that  $\nu \nabla^8 q$  filters out energy buildup on small scales when finite discretization is enforced. We numerically generate the true signal by resolving (2.12) with  $128 \times 64 \times 2$  Fourier modes, which correspond to  $128 \times 128 \times 2$  grid points.

There are two important nondimensional parameters associated with this model [24]

$$b = \frac{\beta}{U_o} \left(\frac{L}{2\pi}\right)^2$$

where  $U_o = 1$  is the horizontal, nondimensionalized velocity scale and  $L = 2\pi$  is the domain size in both directions, and

$$F = \frac{1}{L_d^2} \left(\frac{L}{2\pi}\right)^2.$$

To imitate a turbulent jet in the ocean, we chose a short Rossby radius such that  $F = L_d^{-2} = 40$ , as in [36]. Therefore, because  $k_d^2 = \sqrt{8}/L_d = 320$  is large, (2.12) is numerically stiff.

The large-scale components of this system are barotropic, while the small-scale components are baroclinic [61, 36]. For the two layer model, the barotropic streamfunction is defined as the average between the two layers,  $\psi = (\psi_1 + \psi_2)/2$ . Similarly, the barotropic QG potential vorticity  $q$  is given by the average of the QG potential vortices in each layer. The baroclinic



terms are defined by  $\psi^c = (\psi_1 - \psi_2)/2$  and  $q^c = (q_1 - q_2)/2$ . Then (2.12) becomes

$$\begin{aligned} \frac{\partial q}{\partial t} + J(\psi, q) &+ \beta \frac{\partial \psi}{\partial x} + \kappa \nabla^8 q + \\ &+ \left[ J(\psi^c, c^c) + U \frac{\partial \nabla^2 \psi^c}{\partial x} - \kappa \nabla^2 \psi^c \right] = 0. \end{aligned} \quad (2.13)$$

We plot the barotropic and baroclinic velocity fields at two times in Figure 2.1. Notice that this system exhibits transitions between zonal (east/west) flow and large scale Rossby waves, which propagate northward and southward [24].

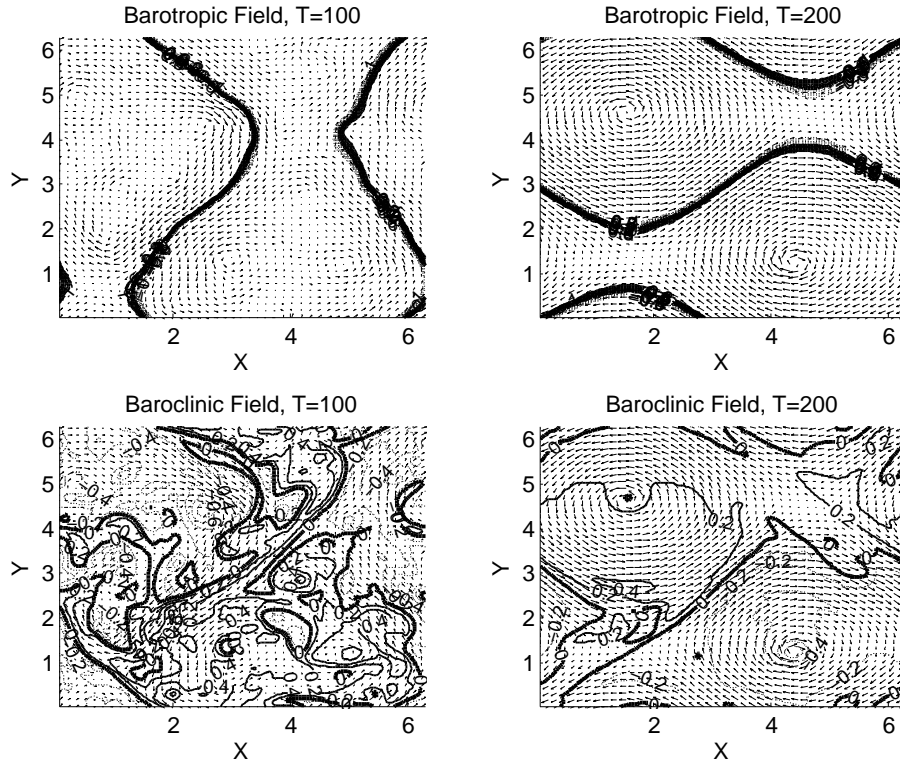


Figure 2.1: The barotropic velocity field (arrows) and the streamfunction,  $\psi$ , (contour) (top) and the baroclinic velocity field and streamfunctions,  $\psi^c$  (bottom) at times  $T = 100$  and  $200$ .

We will consider  $M$  randomly located observations of the large-scale (barotropic) streamfunction corrupted by an i.i.d Gaussian noise with mean zero and variance  $r^o$ , chosen to correspond to approximately 10% and 25% of the climatological barotropic streamfunction variance,

$E = \text{Var}(\psi) \approx 69.46$ . In reality, observations of the small-scale (baroclinic) streamfunctions are not usually available in the ocean, while observations of the large-scale (barotropic) streamfunctions are sparse. This setup makes this test problem difficult for filtering, as the baroclinic terms within the brackets of (2.13) create a numerical stiffness.

## 2.2 Hierarchical Bayesian Filtering Approach

The proposed hierarchical Bayesian filtering approach consists of two steps: the first step is to interpolate the irregularly spaced observations  $\tilde{v}$  to  $v$  through an appropriate interpolation scheme and obtain  $P(v|u, \tilde{v})$  in (2.3); the second step is to apply  $P(u)P(v|u, \tilde{v})$  through a reduced stochastic filtering algorithm. Here we will compare two different interpolations in the first step: a deterministic linear interpolation and a geostatistical interpolation technique called ordinary kriging.

Let  $\tilde{s}_i = (\tilde{x}_i, \tilde{y}_i)$  for  $i = 1, \dots, M$  denote the irregularly spaced observation locations on a  $[0, 2\pi] \times [0, 2\pi]$  periodic domain and let  $s_{i,j} = (x_i, y_i)$ ,  $i, j = 1, \dots, N$  with

$$x_i = C + \frac{2\pi i h}{N}, \quad y_j = C + \frac{2\pi j h}{N},$$

for constant  $C$  and distance  $h$  denote the regularly spaced grid points. In our numerical experiments, we interpolate to  $N = 6$  grid points on each axis. In Sections 2.2.1 and 2.2.2, we discuss two techniques that interpolate  $Z(\tilde{s}_i)$ , the noisy barotropic streamfunction,

$$Z(\tilde{s}_i) = \psi(\tilde{s}_i) + \tilde{\sigma}_i^o, \quad \tilde{\sigma}_i^o \sim \mathcal{N}(0, r^o), \quad (2.14)$$

where  $\psi$  is a solution of (2.13), into  $Z(s_{i,j})$ . We then discuss the reduced filtering approach in Section 2.2.3.

### 2.2.1 Linear Interpolation

There are non-unique criteria to interpolate two-dimensional data from irregularly spaced observation networks. For simplicity, we choose to use just three observations to interpolate at each regular grid point. These three are selected so that they form a triangle with the grid point in their interior and so that their distances from the interpolation grid point are minimal. Since these criteria are arbitrary, we expect that the linear interpolation results in Section (2.3) will not be robust when the criteria are changed.

Denote the three irregularly spaced nodes  $\tilde{s}_m$ ,  $\tilde{s}_n$ , and  $\tilde{s}_p$ , and suppose that we interpolate

$Z(s_{i,j})$ . Then

$$Z(x_i, y_j) = L_m(x_i, y_j)Z(\tilde{s}_m) + L_n(x_i, y_j)Z(\tilde{s}_n) + L_p(x_i, y_j)Z(\tilde{s}_p),$$

where

$$L_m(x, y) = \frac{(\tilde{y}_n - y)(\tilde{x}_n - \tilde{x}_p) - (\tilde{x}_n - x)(\tilde{y}_n - \tilde{y}_p)}{(\tilde{y}_n - \tilde{y}_m)(\tilde{x}_n - \tilde{x}_p) - (\tilde{x}_n - \tilde{x}_m)(\tilde{y}_n - \tilde{y}_p)}$$

is a Lagrange function about three points. The other Lagrange functions,  $L_n(x, y)$  and  $L_p(x, y)$  are defined in the same fashion.

In the right panel of Figure 2.2, we show an example linear interpolation of  $M = 36$  irregularly spaced, sparse observations, with locations depicted by circles, at time  $T = 340$  time units. For comparison, we include the truth at this time, resolved at 36 coarse grid points. There is a relatively good agreement between the interpolated observations and the truth.

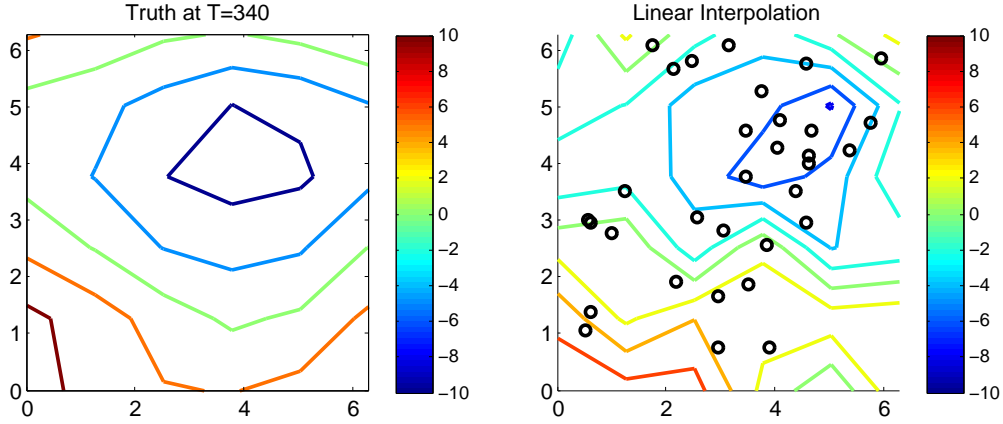


Figure 2.2: True process, compared to a linear interpolation at  $T = 340$ , with  $r^o = 17.3$ . The circles give observation locations and the contours give the barotropic streamfunction,  $\psi$

To quantify the uncertainty in this method, we compute the error covariance matrix in Fourier space. For an interpolated observation  $z_{j,m} = Z(x_j, y_m)$ , where  $x_j = jh$ ,  $y_m = mh$ , and  $Nh = 2\pi$  for distance  $h$ , let us denote the three corresponding observations as  $\tilde{z}_{j,m,1}$ ,  $\tilde{z}_{j,m,2}$ , and  $\tilde{z}_{j,m,3}$ . To simplify the notation, let us also denote  $C_{j,m,s} = L_s(x_j, y_m)$ . Thus

$$z_{j,m} = C_{j,m,1}\tilde{z}_{j,m,1} + C_{j,m,2}\tilde{z}_{j,m,2} + C_{j,m,3}\tilde{z}_{j,m,3}.$$

The discrete Fourier transform of  $z_{j,m}$  is then given by

$$\begin{aligned}\hat{z}_{k,l} &= \frac{h^2}{4\pi^2} \sum_{j=1}^N \sum_{m=1}^N z_{j,m} e^{-ikjh} e^{-ilmh} \\ &= \frac{h^2}{4\pi^2} \sum_{j=1}^N \sum_{m=1}^N \sum_{s=1}^3 C_{j,m,s} \tilde{z}_{j,m,s} e^{-i(kj+lm)h}.\end{aligned}$$

Additionally, let  $\hat{\sigma}_{k,l}^o$  denote the  $(k,l)$ -Fourier component of the interpolated noise,  $\sigma_{j,m}$ , and let  $\{\tilde{\sigma}_{j,m,s}\}_{s=1,2,3}$  denote the noises of the three irregularly spaced observations associated with  $z_{j,m}$ . Then the cross-covariance of the noise between wavenumbers  $(k,l)$  and  $(k',l')$  is given in [8] by

$$\begin{aligned}\langle \hat{\sigma}_{k,l}^o (\hat{\sigma}_{k',l'}^o)^* \rangle &= \left\langle \left( \frac{h^2}{4\pi^2} \sum_{j=1}^N \sum_{m=1}^N \sum_{s=1}^3 C_{j,m,s} \tilde{\sigma}_{j,m,s} e^{i(kj+lm)h} \right) \times \right. \\ &\quad \left. \left( \frac{h^2}{4\pi^2} \sum_{j'=1}^N \sum_{m'=1}^N \sum_{s'=1}^3 C_{j',m',s'} \tilde{\sigma}_{j',m',s'} e^{i(k'j'+l'm')h} \right) \right\rangle \\ &= \left\langle \frac{h^4}{16\pi^4} \left( \sum_{j,m=1}^N \sum_{s=1}^3 C_{j,m,s} \tilde{\sigma}_{j,m,s} e^{-i(kj+lm)h} \right) \times \right. \\ &\quad \left. \left( \sum_{j',m'=1}^N \sum_{s'=1}^3 C_{j',m',s'} \tilde{\sigma}_{j',m',s'} e^{-i(k'j'+l'm')h} \right) \right\rangle,\end{aligned}$$

where  $\langle \bullet \rangle$  denotes expectation with respect to the Gaussian noise distribution. The diagonal terms, for which  $k = k'$  and  $l = l'$ , simplify to

$$\begin{aligned}\langle |\sigma_{k,l}^o| \rangle &= \frac{r^o h^4}{16\pi^4} \sum_{j,m=1}^N \sum_{j',m'=1}^N \sum_{s,s'=1}^3 C_{j,m,s} C_{j',m',s'} e^{-i[k(j-j')+l(m-m')]h} \\ &= \frac{r^o h^4}{16\pi^4} \left( \sum_{j,m=1}^N \sum_{s=1}^3 C_{j,m,s}^2 + \sum_{j,m=1}^N \sum_{j',m'=1}^N \sum_{s,s'=1}^3 C_{j,m,s} C_{j',m',s'} e^{-i[k(j-j')+l(m-m')]h} \right),\end{aligned}$$

where  $r^o = \langle \tilde{\sigma}_{j,m,s} \tilde{\sigma}_{j',m',s'} \rangle$ . The  $C_{j,m,s}^2$  term in the second line corresponds to the contributions of the diagonal terms of the physical space covariance matrix, while the other terms correspond to the contributions of the off-diagonal terms of the physical space covariance.

In physical space, the noise covariance matrix, illustrated in the left panel of Figure 2.3, is diagonally dominant with nearby nonzero correlations as expected, due to the interpolation. Observe that the error covariance matrix depends on observation locations, rather than

the observations themselves. In Fourier space, the error covariance matrix of the interpolated observations is diagonally dominant (see the right panel of Figure 2.3).

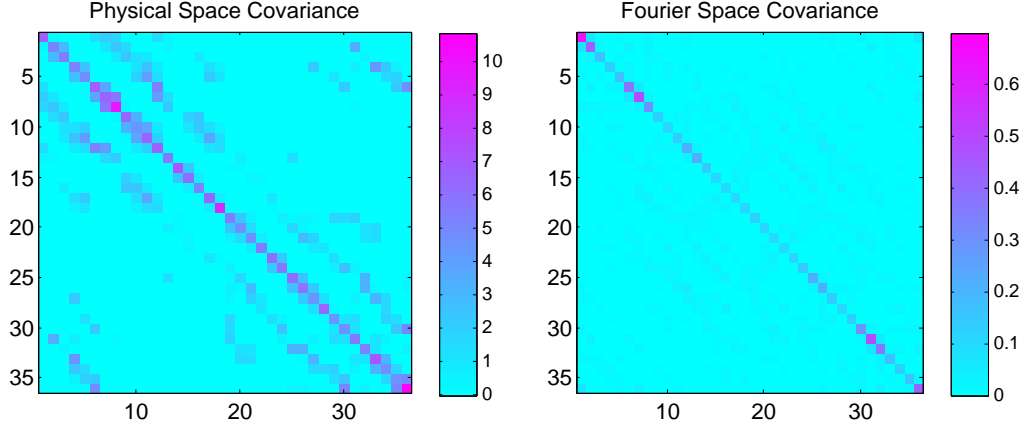


Figure 2.3: The noise covariance matrices associated with linear interpolation at  $T = 340$ , with  $r^o = 17.3$  in physical space (left) Fourier space (real components, right). The imaginary component is one magnitude order smaller (not shown).

### 2.2.2 Ordinary Kriging

Kriging is a maximum likelihood estimator of a random field  $Z$  modeled by  $Z(s) = \mu(s) + \delta(s)$ , with Gaussian, stationary noises  $\delta(s) \sim \mathcal{N}(0, C(s, s))$ . The key idea in kriging is to fit the observations

$$\vec{Z} = [Z(\tilde{s}_1), Z(\tilde{s}_2), \dots, Z(\tilde{s}_M)]^T$$

to an empirically chosen isotropic parametric covariance function  $C$ . By isotropic, we mean that the covariance depends on the distance between observations, that is,  $C(s_1, s_2) = C(\|s_1 - s_2\|)$ .

We will choose  $C$  from an exponential family with two parameters, and use a nonlinear optimization algorithm to obtain the parameters. These parameters change when different sets of observations become available at different times. In this sense, kriging is a data-driven interpolation scheme, even when the observation network is fixed. To the contrary, the Lagrange functions associated with linear interpolation are constant for a fixed observation network.

To model the covariance function  $C$ , let us first define a variogram as a functional  $2\gamma$  that

satisfies the relation

$$2\gamma(\tilde{s}_i - \tilde{s}_j) \equiv \text{Var}(Z(\tilde{s}_i) - Z(\tilde{s}_j)) = \langle [Z(\tilde{s}_i) - Z(\tilde{s}_j)]^2 \rangle, \quad (2.15)$$

to characterize the spatial dependence of the process  $Z$  [10]. Expanding the right hand side of (2.15), we obtain

$$\begin{aligned} 2\gamma(\tilde{s}_i - \tilde{s}_j) &= \text{Var}(Z(\tilde{s}_i)) + \text{Var}(Z(\tilde{s}_j)) - 2\text{Cov}(Z(\tilde{s}_i), Z(\tilde{s}_j)) \\ &= 2(r^o - C(\|\tilde{s}_i - \tilde{s}_j\|)), \end{aligned}$$

where we use  $r^o = C(0)$  and the stationary assumption. This yields a convenient relation between covariance and the variogram:

$$C(\tilde{s}_i, \tilde{s}_j) = C(\|\tilde{s}_i - \tilde{s}_j\|) = r^o - \gamma(\tilde{s}_i - \tilde{s}_j). \quad (2.16)$$

While the exact variogram cannot be known, we can estimate it by using the irregularly spaced observations. To build a variogram estimator, denoted  $\hat{\gamma}(r)$ , we follow the description in [10]. We first place the observations into bins based on distance from one another. A bin consists of all pairs of observations which are within distance  $r$  of each other, where  $r$  is a small distance. Then for each bin, we compute

$$2\hat{\gamma}(r) = \frac{1}{|N(r)|} \sum_{i,j \in N(r)} (R(\tilde{s}_i) - R(\tilde{s}_j))^2,$$

where  $R(\tilde{s}_i) = Z(\tilde{s}_i) - \mu(\tilde{s}_i)$  is a residual and  $N(r)$  is the size of the bin. This equation defines the variogram estimator. By computing  $\hat{\gamma}(r)$  for several values of  $r$  we obtain an estimate of the covariance structure in (2.16). Because  $\hat{\gamma}(r)$  depends on the observations, kriging is data driven.

To determine the residual  $R(\tilde{s}_i)$ , we must estimate the mean  $\mu(\tilde{s}_i)$ . Various types of kriging model the mean  $\mu(s)$  differently. We consider ordinary kriging, which assumes that  $\mu(s)$  is locally constant, as is consistent with the stationary assumption. In our numerical experiments, we estimate  $\mu(\tilde{s}_i)$  with median polishing, an algorithm that estimates the mean value at an observation location by averaging the observations in the same mesh row and column.

We construct bins for several values of  $r$  so that we can estimate  $\gamma(\tilde{s}_i - \tilde{s}_j)$  in (2.16) for numerous values of  $\|\tilde{s}_i - \tilde{s}_j\|$ . Based on the shape of the variogram estimator as a function of  $r$ , we choose an appropriate parametric function  $\hat{\gamma}^*(r)$  to model  $\hat{\gamma}(r)$ . An example plot of a variogram fit at time  $T = 340$  time units is given in Figure 2.4. The values of  $\hat{\gamma}(r)$ , depicted as circles in Figure 2.4, appear to follow an exponential curve. Therefore, we choose  $\gamma$  from an

exponential family,

$$\hat{\gamma}^*(r) = \sigma^2 \exp(-\rho r), \quad r \geq 0, \quad (2.17)$$

where  $\rho, \sigma > 0$  are the parameters to be determined. Note that there are many other parametric forms that can be used beside the exponential family [10]. The appropriate form is dictated by the data. The curve in Figure 2.4 gives the least squares fit of (2.17) with respect to  $\rho$  and  $\sigma$  to  $\hat{\gamma}(r)$  at time  $T = 340$ . The corresponding parameters of the residual process at this time are  $\sigma = 0.73254$  and  $\rho = 1.3019$ . The covariance is estimated by

$$C(\tilde{s}_i, \tilde{s}_j) = r^o - \sigma^2 \exp(-\rho \|\tilde{s}_i - \tilde{s}_j\|).$$

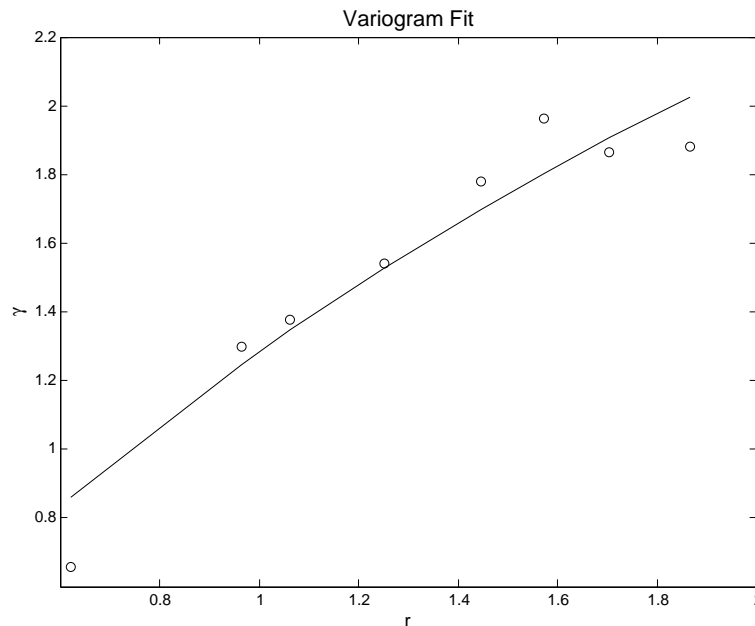


Figure 2.4: An exponential variogram fit at  $T = 340$ , with  $r^o = 17.3$ , plotted as a function of  $r$ .

The estimator for  $Z$  at location  $s_{i,j}$  is given by the conditional expectation  $\mathbb{E}(Z(s_{i,j})|\vec{Z})$ . The uncertainty of this estimator is given by the conditional covariance,  $Cov(Z(s_{i,j}), Z(s_{i',j'})|\vec{Z})$ . Given the observations

$$\vec{Z} = [Z(\tilde{s}_1), Z(\tilde{s}_1), \dots, Z(\tilde{s}_M)]^T,$$

the joint distribution  $(Z(s_{i,j}), \vec{Z})$  is given by

$$(Z(s_{i,j}), \vec{Z}) \sim \mathcal{N} \left( (\mu(s_{i,j}), \vec{\mu}), \begin{bmatrix} C(s_{i,j}, s_{i,j}) & c(s_{i,j})^T \\ c(s_{i,j}) & \Sigma \end{bmatrix} \right),$$

where

$$c(s_{i,j}) = [C(s_{i,j}, \tilde{s}_1), C(s_{i,j}, \tilde{s}_2), \dots, C(s_{i,j}, \tilde{s}_M)], \quad \Sigma_{i,j} = C(\tilde{s}_i, \tilde{s}_j).$$

In general, kriging gives the conditional expectation  $\mathbb{E}(Z(s_{i,j})|\vec{Z})$  by

$$\begin{aligned} \mathbb{E}(Z(s_{i,j})|\vec{Z}) &\equiv \mu(s_{i,j}) + c(s_{i,j})^T \Sigma^{-1} (\vec{Z} - \vec{\mu}) \\ &= c(s_{i,j})^T \Sigma^{-1} \vec{Z} + \mu(1 - c(s_{i,j})^T \Sigma^{-1} \vec{1}) \\ &= \vec{\lambda}^T \vec{Z} + \mu(1 - \vec{\lambda}^T \vec{1}), \end{aligned}$$

where  $\vec{\lambda} = \Sigma^{-1} c(s_{i,j})$ . By the stationary assumption, we assume that, locally,  $\mu(s_{i,j}) = \mu(\tilde{s}_k) = \mu$ ; here  $\mu(\tilde{s}_k)$  is a component of  $\vec{\mu}$ . See Theorem 3.5 of [30] for the detailed derivation of conditional expectation of a multivariable Gaussian distribution. Ordinary kriging requires that  $\vec{\lambda}^T \vec{1} = 1$ , which eliminates the dependency of the estimator on the unknown locally constant  $\mu(s)$ .

Ordinary kriging is mathematically defined with the estimator

$$\mathbb{E}(Z(s_{i,j})|\vec{Z}) = \sum_{k=1}^M \lambda_{i,j,k} Z(\tilde{s}_k) \tag{2.18}$$

such that the mean square predicted error,

$$MSPE(\vec{\lambda}) = \mathbb{E} \left[ \left( Z(s_{i,j}) - \sum_{k=1}^M \lambda_{i,j,k} Z(\tilde{s}_k) \right)^2 \right], \tag{2.19}$$

is minimized, subject to the constraint

$$\sum_{k=1}^M \lambda_{i,j,k} = 1.$$

Rewrite the sum in (2.19) so that

$$MSPE(\vec{\lambda}) = Var \left( \sum_{k=0}^M a_k Z_k \right)$$



where  $a_0 = 1$ ,  $a_k = -\lambda_{i,j,k} = -\lambda_k$  for  $k > 0$ ,  $Z_0 = Z(s_{i,j})$  and  $Z_k = Z(\tilde{s}_k)$  for  $k > 0$ . Then

$$MSPE(\vec{\lambda}) = C(Z_0, Z_0) - 2 \sum_{i=0}^M \lambda_i C(Z_i, Z_0) + \sum_{i,j=1}^M \lambda_i \lambda_j C(Z_i, Z_j).$$

The constraint minimization problem in (2.19) is equivalent to minimizing

$$MSPE(\vec{\lambda}) + 2\alpha \left( \sum_{k=1}^M \lambda_k - 1 \right) \equiv L(\vec{\lambda}, \alpha).$$

By setting

$$\frac{\partial L}{\partial \lambda_k} = 0, \quad \text{and} \quad \frac{\partial L}{\partial \alpha} = 0$$

we obtain

$$-2C(Z_0, Z_k) + 2 \sum_{j=1}^M K \lambda_j C(Z_j, Z_k) + 2\alpha = 0$$

and

$$\sum_{k=1}^M \lambda_k - 1 = 0.$$

This can be written as a linear system  $\Gamma_0 \vec{\lambda}_0 = \vec{\gamma}_0$  with

$$\Gamma_0 = \left[ \begin{array}{ccc|c} & & & 1 \\ & C(\tilde{s}_k, \tilde{s}_l) & & \vdots \\ & & & 1 \\ \hline 1 & \dots & 1 & 0 \end{array} \right], \quad \vec{\lambda}_0 = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_M \\ \alpha \end{bmatrix}, \quad \vec{\gamma}_0 = \begin{bmatrix} C(s_{i,j}, \tilde{s}_1) \\ \vdots \\ C(s_{i,j}, \tilde{s}_n) \\ 1 \end{bmatrix}$$

In our notation, the multipliers are thus given by

$$\begin{bmatrix} \lambda_{i,j,1} \\ \vdots \\ \lambda_{i,j,n} \\ \alpha \end{bmatrix} = \left[ \begin{array}{ccc|c} & & & 1 \\ & C(\tilde{s}_k, \tilde{s}_l) & & \vdots \\ & & & 1 \\ \hline 1 & \dots & 1 & 0 \end{array} \right]^{-1} \begin{bmatrix} C(s_{i,j}, \tilde{s}_1) \\ \vdots \\ C(s_{i,j}, \tilde{s}_n) \\ 1 \end{bmatrix}. \quad (2.20)$$

The conditional variance is then given as

$$\begin{aligned}
MSPE(\vec{\lambda}_0) &= C(Z_0, Z_0) - 2 \sum_{i=1}^M \lambda_i C(Z_i, Z_0) + \sum_{i=1}^M \sum_{j=1}^M \lambda_i \lambda_j C(Z_i, Z_j) \\
&= C(Z_0, Z_0) - 2 \sum_{i=1}^M \lambda_i C(Z_i, Z_0) + \sum_{i=1}^M \lambda_i [C(Z_i, Z_0) - \alpha] \\
&= C(Z_0, Z_0) - \sum_{i=1}^M \lambda_i C(Z_i, Z_0) - \alpha \\
&= C(Z_0, Z_0) - \vec{\lambda}_0^T \vec{\gamma}_0.
\end{aligned}$$

Then

$$var[Z_0|\vec{Z}] = C(Z_0, Z_0) - \vec{\gamma}_0^T \Gamma_0^{-1} \vec{\gamma}_0$$

and

$$var[Z(s_{i,j})|\vec{Z}] = C(s_{i,j}, s_{i,j}) - [c(s_{i,j})^T, 1] \Gamma_0^{-1} [c(s_{i,j})^T, 1]^T.$$

Thus the generalized covariance is

$$cov[Z(s_{i,j})Z(s_{i',j'})|\vec{Z}] = C(s_{i,j}, s_{i',j'}) - [c(s_{i,j}), 1] \Gamma_0^{-1} [c(s_{i',j'})^T, 1]^T.$$

In the right panel of Figure 2.5, we show a kriging result at time  $T = 340$ . We again include the truth for comparison. We again observe a good agreement between the truth and the reconstructed field.

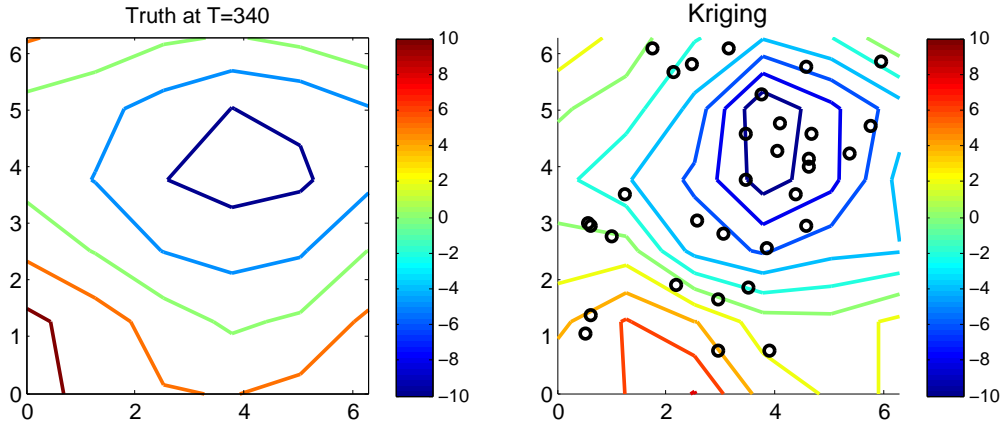


Figure 2.5: True process, compared with kriging at  $T = 340$ , with  $r^o = 17.3$ . The circles give observation locations and the contours give the barotropic streamfunction,  $\psi$ .

Figure 2.6 shows the associated error covariance matrix for the kriging result in both physical space and Fourier space. By design, the error covariance matrix in Fourier space is diagonal because of the stationarity of the covariance estimator with isotropic function  $C$ . The slight nonzero non-diagonal terms in the Fourier domain in Figure 2.6 are numerical artifacts. In contrast to the linear interpolation scheme, the error covariance matrix obtained from kriging changes in time even when observation locations do not, because  $C(\tilde{s}_i, \tilde{s}_j)$  in (2.16) depends on a variogram estimator that changes based on the observations.

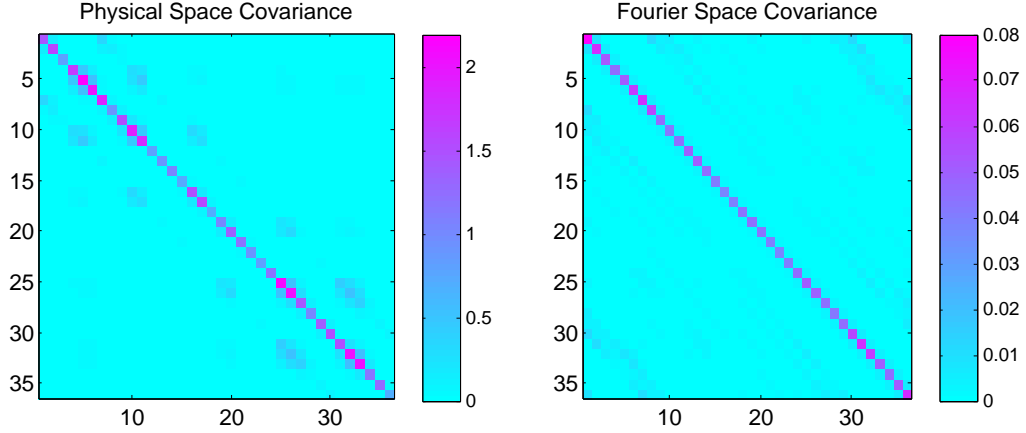


Figure 2.6: The noise covariance matrices associated with kriging at  $T = 340$ , with  $r^o = 17.3$  in physical space (left) and Fourier space (real components, right). The imaginary component is one magnitude order smaller (not shown).

### 2.2.3 Reduced Stochastic Filter

The second step in the hierarchical Bayesian framework in (2.3) is to apply the reduced stochastic filter. We will report the numerical simulations results using a simple reduced stochastic model, the Mean Stochastic Model (MSM) [26, 48]. In this method, we fit the variance and the integral of the autocorrelation time function for each mode to the empirical solutions of the QG model. We will construct the MSM in Fourier space for the barotropic mode dynamics in (2.13). Following standard closure modeling for turbulent systems [61, 50, 51, 43, 11, 47], we replace the baroclinic terms and the nonlinear advective terms in (2.13) with stochastic noise and linear damping terms, such that each Fourier component can be written as

$$d\hat{\psi}_{k,l}(t) = (-d_{k,l} + i\omega_{k,l})\hat{\psi}_{k,l}(t)dt + \bar{\hat{\psi}}_{k,l}dt + \sigma_{k,l}dW_{k,l}(t), \quad (2.21)$$

where  $\hat{\psi}_{k,l}$  denotes the Fourier component of the barotropic streamfunction  $\psi$  and  $W_{k,l}(t) = (W_1(t) + iW_2(t))/\sqrt{2}$  denotes a complex-valued Wiener process with independent standard Wiener processes,  $W_1$  and  $W_2$ . Each component of  $W_{k,l}$  satisfies  $dW_j(t) = \dot{W}_j(t)dt$ , where  $\langle \dot{W}_j \rangle = 0$ . Therefore, white noise is a derivative of the Wiener process. The exact solution of (2.21) is given by

$$\hat{\psi}_{k,l}(t) = \hat{\psi}_{k,l}(0)e^{(-d_{k,l}+i\omega_{k,l})t} + \sigma_{k,l} \int_0^t e^{(-d_{k,l}+i\omega_{k,l})(t-\tau)} dW_{k,l}(\tau).$$

For simplicity, we assume that  $\hat{\psi}_{k,l}$  has an initial variance of zero. The mean of  $\hat{\psi}_{k,l}(t)$  is given by

$$\begin{aligned} \bar{\hat{\psi}}_{k,l} &= \langle \hat{\psi}_{k,l}(t) \rangle \\ &= \left\langle \hat{\psi}_{k,l}(0)e^{(-d_{k,l}+i\omega_{k,l})t} + \sigma_{k,l} \int_0^t e^{(-d_{k,l}+i\omega_{k,l})(t-\tau)} dW_{k,l}(\tau) \right\rangle \\ &= e^{(-d_{k,l}+i\omega_{k,l})t} \langle \hat{\psi}_{k,l}(0) \rangle + \sigma_{k,l} \int_0^t e^{(-d_{k,l}+i\omega_{k,l})(t-\tau)} \langle dW_{k,l}(\tau) \rangle \\ &= e^{(-d_{k,l}+i\omega_{k,l})t} \langle \hat{\psi}_{k,l}(0) \rangle + \sigma_{k,l} \sum_{j=1}^2 \int_0^t e^{(-d_{k,l}+i\omega_{k,l})(t-\tau)} \langle \dot{W}_j(\tau) \rangle d\tau \\ &= e^{(-d_{k,l}+i\omega_{k,l})t} \langle \hat{\psi}_{k,l}(0) \rangle. \end{aligned}$$

The variance of  $\hat{\psi}_{k,l}(t)$  is given by

$$\begin{aligned} \text{Var}[\hat{\psi}_{k,l}(t)] &= \langle (\hat{\psi}_{k,l}(t) - \bar{\hat{\psi}})(\hat{\psi}_{k,l}(t) - \bar{\hat{\psi}})^* \rangle \\ &= \left\langle \sigma_{k,l}^2 \int_0^t e^{(-d_{k,l}+i\omega_{k,l})(t-\tau')} dW_{k,l}(\tau') \left( \sigma_{k,l} \int_0^t e^{(-d_{k,l}+i\omega_{k,l})(t-\tau)} dW_{k,l}(\tau) \right)^* \right\rangle \\ &= \frac{\sigma_{k,l}}{2} e^{-2d_{k,l}t} \int_0^t \int_0^t e^{d_{k,l}(\tau'+\tau)-i\omega(\tau'-\tau)} \langle dW_{k,l}(\tau') dW_{k,l}(\tau) \rangle. \end{aligned}$$

The components of  $W_{k,l}$  satisfies the properties  $\langle \dot{W}_1(\tau') \dot{W}_2(\tau) \rangle = \langle \dot{W}_2(\tau') \dot{W}_1(\tau) \rangle = 0$  and  $\langle \dot{W}_1(\tau') \dot{W}_1(\tau) \rangle = \langle \dot{W}_2(\tau') \dot{W}_2(\tau) \rangle = \delta(\tau' - \tau)$ , where  $\delta(x)$ , the Dirac delta function, is defined

such that  $\delta(0) = \infty$  and  $\int_{\mathbb{R}} \delta(x) dx = 1$ . Then

$$\begin{aligned}
\text{Var}[\hat{\psi}_{k,l}(t)] &= \sum_{j=1}^2 \frac{\sigma_{k,l}}{2} e^{-2d_{k,l}t} \int_0^t \int_0^t e^{d_{k,l}(\tau'+\tau)-i\omega(\tau'-\tau)} \left\langle \dot{W}_j(\tau') \dot{W}_j(\tau) \right\rangle d\tau' d\tau \\
&= \sum_{j=1}^2 \frac{\sigma_{k,l}}{2} e^{-2d_{k,l}t} \int_0^t \int_0^t e^{d_{k,l}(\tau'+\tau)-i\omega(\tau'-\tau)} \delta(\tau' - \tau) d\tau' d\tau \\
&= \sigma_{k,l}^2 e^{-2d_{k,l}t} \int_0^t e^{2d_{k,l}\tau} d\tau \\
&= \frac{\sigma_{k,l}^2}{2d_{k,l}} (1 - e^{-2d_{k,l}t}).
\end{aligned}$$

As  $t$  approaches  $\infty$ , the variance converges to the climatological variance of the forced Ornstein-Uhlenbeck process,

$$E_{k,l} = \frac{\sigma_{k,l}^2}{2d_{k,l}}. \quad (2.22)$$

The autocorrelation time of the forced Ornstein-Uhlenbeck process, is given by the limit

$$T_{k,l} + i\theta_{k,l} = \lim_{s,t \rightarrow \infty} \mathcal{E}(s, t),$$

where  $\mathcal{E}(s, t)$  is given by

$$\mathcal{E}(s, t) = \int_0^s \langle (\hat{\psi}_{k,l}(t+u) - \bar{\hat{\psi}}_{k,l}) E_{k,l}^{-1} (\hat{\psi}_{k,l}(u) - \bar{\hat{\psi}}_{k,l})^* \rangle du.$$

We evaluate this expression.

$$\begin{aligned}
\mathcal{E}(s, t) &= \int_0^s \langle (\hat{\psi}_{k,l}(t+u) - \bar{\hat{\psi}}_{k,l}) E_{k,l}^{-1} (\hat{\psi}_{k,l}(u) - \bar{\hat{\psi}}_{k,l})^* \rangle du \\
&= 2d_{k,l} \int_0^s \left\langle \int_0^{t+u} e^{(-d_{k,l} + i\omega_{k,l})(t+u-\tau)} dW(\tau) \int_0^u e^{(-d_{k,l} - i\omega_{k,l})(u-\tau')} dW(\tau') \right\rangle du \\
&= 2d_{k,l} e^{(-d_{k,l} + i\omega_{k,l})t} \int_0^s e^{-2d_{k,l}u} \left[ \int_0^u \int_0^{t+u} e^{d_{k,l}(\tau'+\tau) - i\omega_{k,l}(\tau-\tau')} \langle dW(\tau') dW(\tau) \rangle \right] du \\
&= 2d_{k,l} e^{(-d_{k,l} + i\omega_{k,l})t} \int_0^s e^{-2d_{k,l}u} \left[ \int_0^u \int_0^{t+u} e^{d_{k,l}(\tau'+\tau) - i\omega_{k,l}(\tau-\tau')} \delta(\tau' - \tau) d\tau d\tau' \right] du \\
&= 2d_{k,l} e^{(-d_{k,l} + i\omega_{k,l})t} \int_0^s e^{-2d_{k,l}u} \left[ \int_0^u e^{2d_{k,l}\tau'} d\tau' \right] du \\
&= e^{(-d_{k,l} + i\omega_{k,l})t} \int_0^s e^{-2d_{k,l}u} \left[ e^{2d_{k,l}u} - 1 \right] du \\
&= e^{(-d_{k,l} + i\omega_{k,l})t} \int_0^s du - e^{(-d_{k,l} + i\omega_{k,l})t} \int_0^s e^{-2d_{k,l}u} du \\
&= e^{(-d_{k,l} + i\omega_{k,l})t} s - \frac{1}{-2d_{k,l}} e^{(-d_{k,l} + i\omega_{k,l})t} [e^{-2d_{k,l}s} - 1]
\end{aligned}$$

Applying the limit, we obtain the correlation time

$$T_{k,l} + i\theta_{k,l} = \frac{1}{d_{k,l} - i\omega_{k,l}}. \quad (2.23)$$

For each mode  $(k, l)$  we must determine the damping  $d_{k,l}$ , frequency  $\omega_{k,l}$ , constant external forcing  $\bar{\hat{\psi}}_{k,l}$ , and noise strength  $\sigma_{k,l}$ . As in [45], we set the constant forcing equal to the time average of its associated Fourier component. To obtain  $d_{k,l}$ ,  $\omega_{k,l}$ , and  $\sigma_{k,l}$ , we fit them to the solutions of the forced Ornstein-Uhlenbeck process given in (2.22) and (2.23), where the climatological energy spectrum,  $E_{k,l}$ , and the real and imaginary components of the correlation time,  $T_{k,l}$  and  $\theta_{k,l}$  respectively, are empirically estimated from solutions of the true model in (2.12). In reality, a true model may not exist, but there is usually a training data set (for example, reanalysis in geophysical applications).

In Figure 2.7, we show the empirical distribution of the climatological energy,  $E_{k,l}$ . The horizontal axis in Figure 2.7 corresponds to the barotropic modes, ordered from the largest to the smallest variance, averaged over a long period of time [25]. The first twelve modes are ordered as  $(k, l) = (1, 0), (0, 1), (1, 1), (-1, 1), (0, 2), (2, 0), (2, 1), (-2, 1), (1, 2), (-1, 2), (-2, 2)$  and  $(2, 2)$ . The large-scale zonal jet modes carry the second and fifth largest variances and the Rossby mode  $(1, 0)$  has the largest variance. The magnitude of the variances of the first two modes is on the same order, which indicates competition between two distinct regimes, zonal jets and Rossby waves. The marginal probability distribution function of the solutions of the first two modes are

shown in Figure 2.8 whereas the remaining marginal pdf's have Gaussian shape (see [25]). These marginal pdfs are generated through bin counting the barotropic streamfunction, centered at 0, such that both panels in Figure 2.8 show a histogram  $d\psi = \hat{\psi} - \langle \hat{\psi} \rangle$ , and encompass solutions of (2.13) up to  $T = 400$  time units [8].

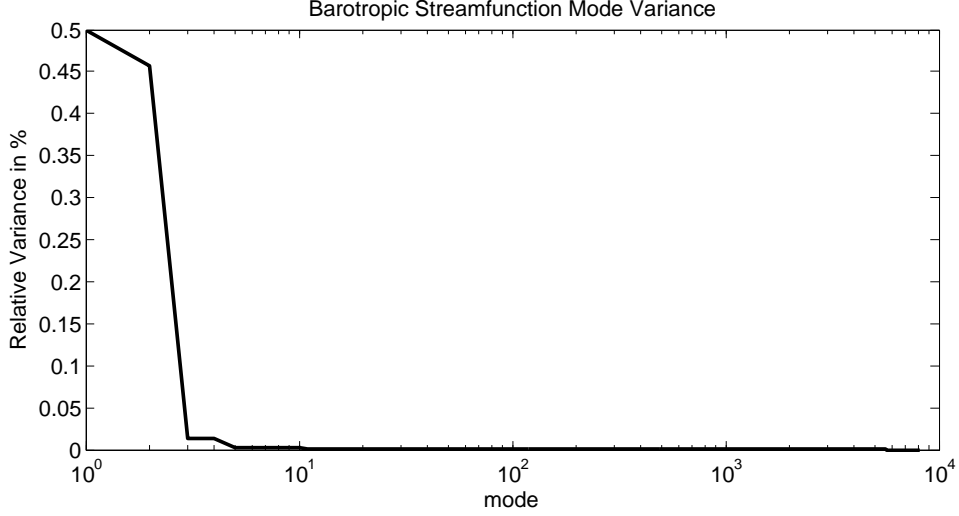


Figure 2.7: Percentage of variances of the barotropic streamfunction as a function of modes, from largest to smallest.

We consider  $M$  irregularly spaced sparse observations of the barotropic streamfunction through the observation model in (2.14). The first step in the hierarchical Bayesian analysis in (2.3) produces the conditional statistics  $\mathbb{E}(Z(s_{i,j})|\vec{Z})$  and  $Cov(Z(s_{i,j})|\vec{Z})$  for  $i, j = 1, \dots, N$ , to be filtered at  $N \times N = 6 \times 6$  modes. With such a coarse resolution, we at least resolve the twelve most energetic modes (see Figure 2.7). In Figure 2.9, we show the decaying time with  $d_{k,l}$ , estimated by solving (2.22)-(2.23); based on this decaying time, we use the observation time intervals  $T_{\text{obs}} = 0.01, 0.04$  and  $0.08$ , which are shorter than the model damping times on the first 12 modes.

The discrete-time reduced filtering model for the second-step in the hierarchical Bayesian analysis in (2.3) is defined as

$$\hat{\psi}_{k,l,m+1} = F_{k,l} \hat{\psi}_{k,l,m} + \tilde{f}_{k,l,m+1} + \eta_{k,l,m+1} \quad (2.24)$$

$$\hat{\psi}_{k,l,m}^o = \tilde{\psi}_{k,l,m} + \epsilon_{k,l,m}^o, \quad (2.25)$$

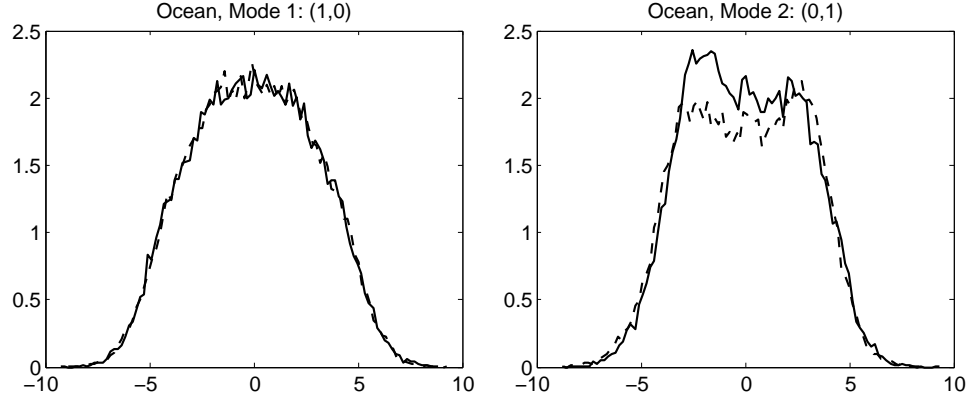


Figure 2.8: Marginal pdfs of the barotropic streamfunction  $\psi$  (centered around their means). Solid lines indicate the real part and dashed lines the imaginary parts.

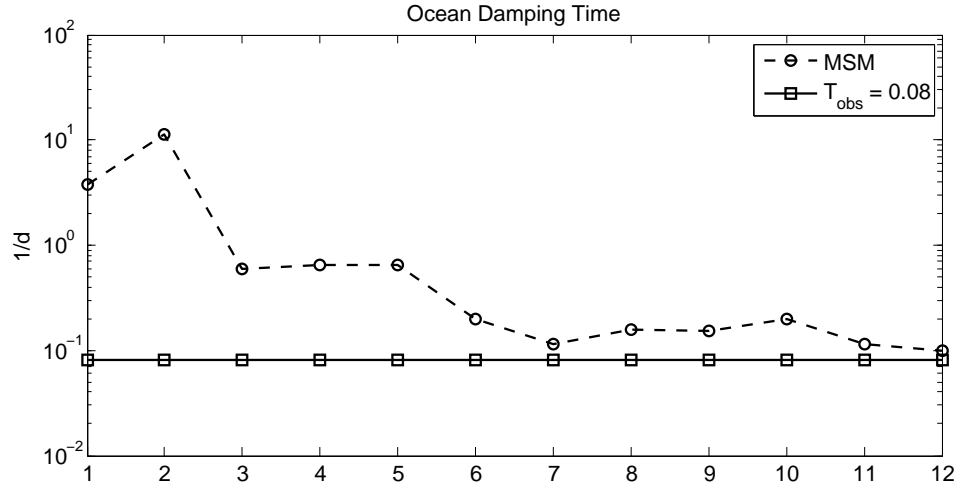


Figure 2.9: Damping time,  $1/d$  of the MSM.

where subscripts  $k, l$  denote the two-dimensional wavenumbers and subscript  $m$  denotes the discrete time step with  $t_{m+1} - t_m = T_{\text{obs}}$ . The parameters in (2.24) are given by exact solution



of the forced Langevin equation in (2.21) with

$$\begin{aligned} F_{k,l} &= e^{(-d_{k,l} + i\omega_{k,l})T_{\text{obs}}}, \\ \tilde{f}_{k,l,m+1} &= \tilde{\psi}_{k,l,m}(1 - F_{k,l}), \\ \eta_{k,l,m+1} &\sim \mathcal{N}(0, Q_{k,l}), \quad Q_{k,l} = \frac{\sigma_{k,l}^2}{2d_{k,l}}(1 - e^{-2d_{k,l}T_{\text{obs}}}). \end{aligned}$$

In (2.25),  $\hat{\psi}_{k,l,m}^o$  denotes the Fourier coefficients of the conditional estimate  $\mathbb{E}(Z_m(s_{i,j})|\vec{Z}_m)$ .

The goal of the Kalman filter is to maximize the conditional distribution

$$P(\hat{\psi}_{k,l,m+1}^a) \sim P(\hat{\psi}_{k,l,m+1}^b)P(\hat{\psi}_{k,l,m+1}^o|\hat{\psi}_{k,l,m+1}^b) = e^{-\frac{1}{2}J(\hat{\psi}_{k,l,m+1}^a)}.$$

This is equivalent to minimizing the cost function

$$J(\hat{\psi}) = \frac{1}{r_{m+1}^b}(\hat{\psi} - \tilde{\psi}_{m+1}^b)^*(\hat{\psi} - \tilde{\psi}_{m+1}^b) + \frac{1}{r_o}(\hat{\psi}_{k,l,m+1}^o - \hat{\psi})^T(\hat{\psi}_{k,l,m+1}^o - \hat{\psi}),$$

where  $\hat{\psi}_{k,l,m+1}^o$  denotes the observation Fourier component.  $J(\hat{\psi})$  attains its minimum for  $\hat{\psi} = \tilde{\psi}_{k,l,m+1}^a$ , which is given by

$$\tilde{\psi}_{k,l,m+1}^a = \tilde{\psi}_{k,l,m+1}^b + K_{k,l,m+1}(\hat{\psi}_{k,l,m+1}^o - \tilde{\psi}_{k,l,m+1}^b),$$

where

$$K_{k,l,m+1} = R_{k,l,m+1}^b(R_{k,l,m+1}^b + \hat{R}_{k,l,m}^o)^{-1}$$

denotes the Kalman gain matrix. The filter completely trusts the model dynamics when  $K_{k,l,m+1} = 0$ . On the other hand, when  $K_{k,l,m+1} = 1$ , the filter trusts the observations.

The prior error covariance is given by

$$R_{k,l,m+1}^b = F_{k,l}R_{k,l,m}^a F_{k,l}^* + Q_{k,l}.$$

When the filter updates with an observation, the new posterior error covariance is obtained by taking the expectation of  $(\hat{\psi}_{k,l,m+1}^b - \tilde{\psi}_{k,l,m+1}^a)(\hat{\psi}_{k,l,m+1}^b - \tilde{\psi}_{k,l,m+1}^a)^*$ . Therefore,

$$\begin{aligned} R_{k,l,m+1}^a &= \left\langle (\hat{\psi}_{k,l,m+1}^b - \tilde{\psi}_{k,l,m+1}^a)(\hat{\psi}_{k,l,m+1}^b - \tilde{\psi}_{k,l,m+1}^a)^* \right\rangle \\ &= (\mathcal{I} - K_{k,l,m+1}) \left\langle (\hat{\psi}_{k,l,m+1}^b - \tilde{\psi}_{k,l,m+1}^b)(\hat{\psi}_{k,l,m+1}^b - \tilde{\psi}_{k,l,m+1}^b)^* \right\rangle (\mathcal{I} - K_{k,l,m+1}^*) \\ &+ K_{k,l,m+1} \hat{R}_{k,l,m}^o K_{k,l,m+1}^* \\ &= (\mathcal{I} - K_{k,l,m+1}) R_{k,l,m+1}^b \end{aligned}$$

The major computational reduction in the proposed filtering approach in (2.24)-(2.25) is through the assumption that different modes are statistically uncorrelated (see the nearly diagonal observation error covariance matrix shown in Figures 2.3 and 2.6). This assumption only holds for appropriate interpolation schemes as investigated in detail in [23] on a simpler set-up. To summarize, the basic Kalman filter solution to (2.24)-(2.25) produces scalar estimates of the mean and covariance prior and posterior to observation  $\bar{\psi}_{k,l,m+1}^o$  through the following recursive equations:

$$\begin{aligned}
\bar{\psi}_{k,l,m+1}^b &= F_{k,l} \bar{\psi}_{k,l,m}^a + \tilde{f}_{k,l,m+1}, \\
R_{k,l,m+1}^b &= F_{k,l} R_{k,l,m}^a F_{k,l}^* + Q_{k,l}, \\
\bar{\psi}_{k,l,m+1}^a &= \bar{\psi}_{k,l,m+1}^b + (\bar{\psi}_{k,l,m+1}^o - \bar{\psi}_{k,l,m+1}^b), \\
R_{k,l,m+1}^a &= (\mathcal{I} - K_{k,l,m+1}) R_{k,l,m}^b, \\
K_{k,l,m+1} &= R_{k,l,m+1}^b (R_{k,l,m+1}^b + \hat{R}_{k,l,m}^o)^{-1}.
\end{aligned} \tag{2.26}$$

## 2.3 Results

We now report the numerical results of applying the hierarchical Bayesian filtering strategy in (2.3) in assimilating irregularly spaced, noisy observations of the barotropic streamfunction in (2.14) to a variety of observation networks, varying the observation error variance,  $r^o$ , as well as the number of observations,  $M$ . The number of assimilation time steps is given by  $\mathcal{T} = T/T_{obs}$ , where we set  $T = 400$  time units. For example, if  $T_{obs} = 0.01$ , the assimilation runs for 40,000 discrete time steps. In each case, we compare filtering skill produced by the Kalman filter in (2.26) in assimilating the interpolated observations obtained either by the linear interpolation or by kriging. We will refer to these two methods as the *filtered linear interpolation scheme* and the *filtered kriging scheme*. To measure the filtering skill, we compute the Root-Mean-Square error,

$$\text{RMSE} = \left( \frac{1}{(\mathcal{T} - T_o)N^2} \sum_{m=T_o}^{\mathcal{T}} \sum_{i,j=1}^N (\bar{\psi}_{i,j,m}^a - \psi_m(s_{i,j}))^2 \right)^{1/2}, \tag{2.27}$$

and the time average physical space pattern correlation,

$$\mathcal{C} = \frac{1}{\mathcal{T} - T_o} \sum_{m=T_o}^{\mathcal{T}} \frac{(\bar{\psi}_m^a)^T \psi_m}{\|\bar{\psi}_m^a\|_2 \|\psi_m\|_2}, \tag{2.28}$$

between the mean posterior state,  $\bar{\psi}_{i,j,m}^a$ , and the true barotropic streamfunction,  $\psi_m(s_{i,j})$ , ignoring the transient time at the first  $T_o = 200$  steps. To check the effectiveness of the second step in the proposed hierarchical Bayesian filtering scheme, we also report the estimates of the

interpolated observations without implementing stochastic recursive filter (that is, only apply either the linear interpolation or kriging). We will refer to these two methods as the *unfiltered linear interpolated scheme* and the *unfiltered kriging scheme*. To demonstrate the robustness of the filtering skill, we show the averages of the RMS error and correlation in (2.27) and (2.28) over 50 different, randomly chosen, irregularly spaced observation networks.

First, we consider observation networks with  $M = N^2 = 36$  and  $r^o = 6.9$ , where  $N$  is the model resolution in each coordinate axis,  $M$  is the number of observations and  $r^o$  is the observational noise variance, taken to be only 10 % of the climatological (temporally averaged) barotropic streamfunction variance,  $E$ , which is approximately 69.64. In Figure 2.10, we show the true barotropic streamfunction and the corresponding filter estimates at time  $T = 340$ , using a randomly chosen observation network (in circles). Comparing the estimates with the truth, we find that the kriging scheme clearly reconstructs the true dynamics better compared to the linear interpolation scheme. In this scenario, ordinary kriging supersedes the linear interpolation scheme; see the first row of Figure 2.14, which displays the RMS error and correlation as functions of observation time interval,  $T_{\text{obs}}$ . The RMS and correlation measures suggest that the estimates are each approximately constant as  $T_{\text{obs}}$  increases. The dynamical constraint produces improved estimates, by only 0.15 in RMS measure and 2 – 3% in correlation, compared to the unfiltered results.

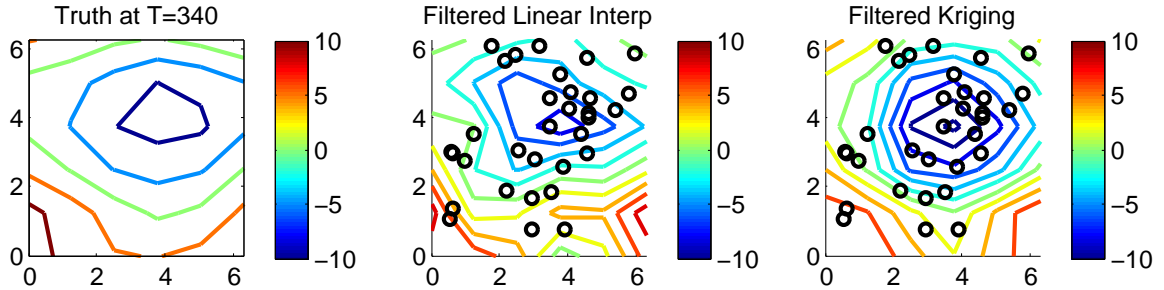


Figure 2.10: The true barotropic streamfunction (left) compared with the filtered linear interpolation scheme (middle) and the filtered kriging scheme (right), with  $M = 36$  and  $r^o = 6.9$  at time  $T = 340$ . The circles depict observation locations.

Next, we consider networks with fewer observations,  $M = 18$ , but subject to the same noise variance  $r^o = 6.9$ . The filtered streamfunction estimates of both schemes are depicted in Figure 2.11. As expected, a sparse observation network deteriorates the accuracy of both methods as compared to the previous case (see also the higher RMS errors and lower correlations in the

second row of Figure 2.14). In this case, the sparse observation networks permit degenerate Lagrange functions for linear interpolation as well as a poor variogram fitting in the covariance estimation in (2.16). Based on the RMS and correlation measures, averaged over 50 different observation networks, we find that both schemes are comparable. For some networks, the linear interpolation scheme can actually provide a better skill relative to kriging, depending on the distribution of the observations. However, the higher filtering skill with linear interpolation in two-dimensions is not robust, since there are non-unique criteria for constructing the Lagrange functions. We also find that the filtered estimates are better than the unfiltered estimates, independent of observational time length. Here, the RMS error and correlation are consistently improved by about 0.2 and 2 – 4%, respectively, when the dynamical constraint is imposed. The numerical result in this scheme suggests that the data-driven step is not effective since the data are too sparse.

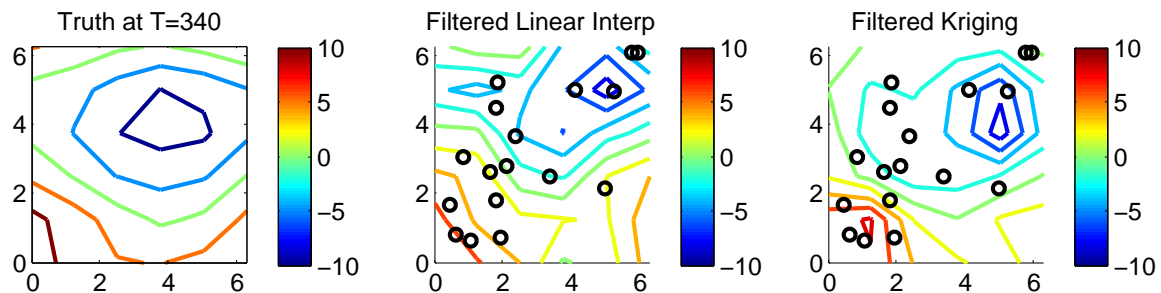


Figure 2.11: The true barotropic streamfunction (left) compared with the filtered linear interpolation scheme (middle) and the filtered kriging scheme (right), with  $M = 18$  and  $r^o = 6.9$  at time  $T = 340$ . The circles depict observation locations.

Now we consider a denser observation network, with more observations than model grid points so that  $M = 49 > N^2 = 36$ , with the same noise variance  $r^o = 6.9$ . From the filtered streamfunction contours in Figure 2.12, it appears that both interpolation schemes produce decent estimates of the true barotropic streamfunction. However, we find that more plentiful observations do not improve the linear interpolation estimates significantly. Comparing the first and third rows in Figure 2.14, we see that the average RMS error decreases from 3.7 to 3.4. The kriging estimates, however, are largely improved in this here; the average RMS error decreases from 3.3 to 2.6. There is a clear improvement in filtering skill relative to the case in which  $M = 36$  above.

Finally, we consider a higher observation noise variance, about  $r^o = 17.3 \approx 25\%E$  in net-

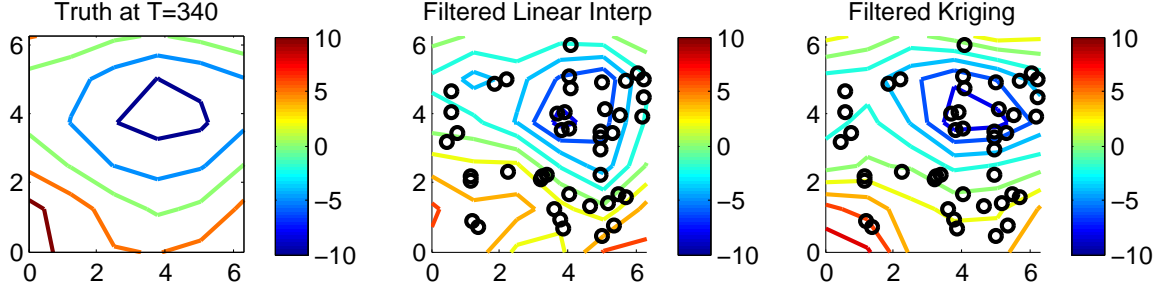


Figure 2.12: The true barotropic streamfunction (left) compared with the filtered linear interpolation scheme (middle) and the filtered kriging scheme (right), with  $M = 49$  and  $r^o = 6.9$  at time  $T = 340$ . The circles depict observation locations.

works with  $M = 36 = N^2$ . From the filtered streamfunction contours, shown in Figure 2.13, it is clear that, once again, the kriging scheme provides better estimates than the linear interpolation scheme. In this case, the average RMS errors of both schemes increase significantly as compared to the first case with smaller noise variance (see the bottom row of Figure 2.14). The poor estimates with the unfiltered linear interpolation and kriging are not surprising, since the data are quite noisy. Notice that the average RMS errors and correlation for the filtered estimates are improved by 0.2 and 3 – 4%, respectively, compared to the unfiltered estimates. This suggests that the dynamical constraint becomes more important when observations are noisy.

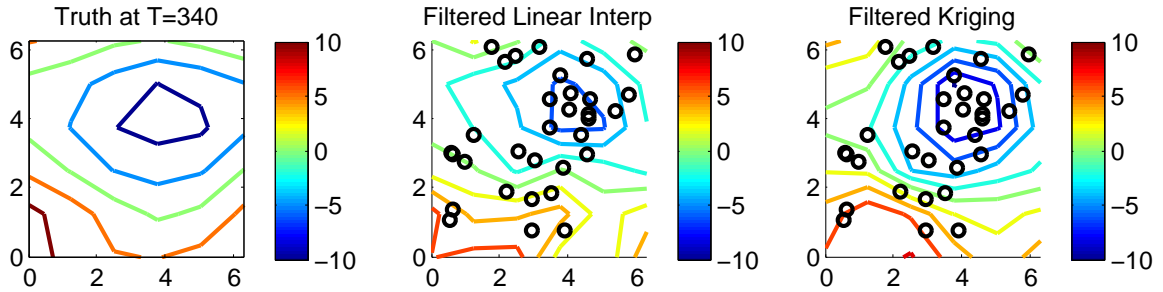


Figure 2.13: The true barotropic streamfunction (left) compared with the filtered linear interpolation scheme (middle) and the filtered kriging scheme (right), with  $M = 36$  and  $r^o = 17.3$  at time  $T = 340$ . The circles depict observation locations.

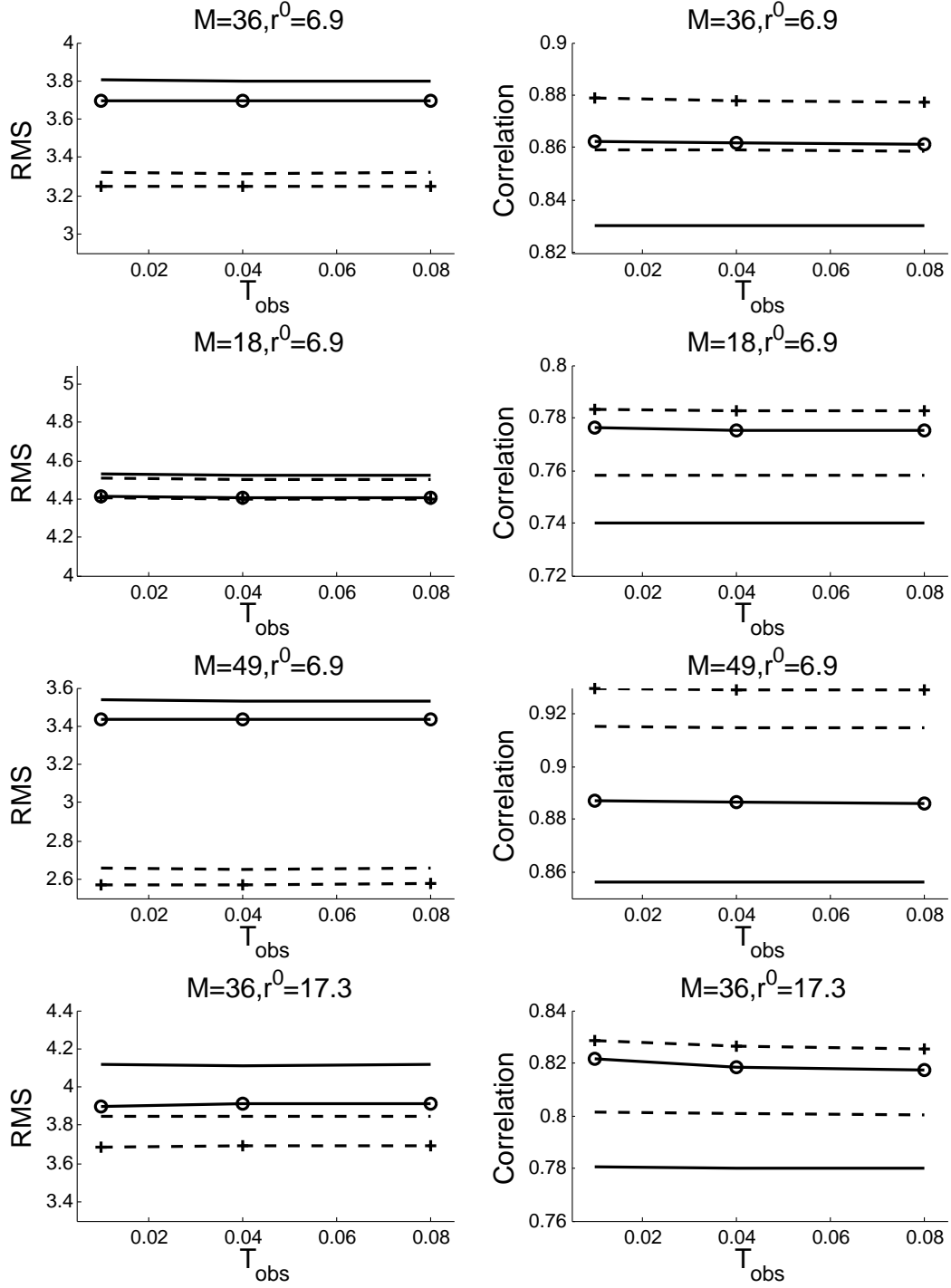


Figure 2.14: The RMS errors (left panels) and correlations (right panels) for each case. Unfiltered kriging (dashes), filtered kriging (dashes with '+' sign), unfiltered linear interpolation (solid line), and filtered linear interpolation (solid line with circles).

For diagnostic purposes, we also report the filtering skill for regularly spaced observations at  $N \times N = 36$  grid points for observations with noise variance  $r^o = 10\%E$  and  $r^o = 25\%$ . Here the filtering skills are improved. In both cases, we observe a lower RMS and corresponding higher correlation with a shorter observational time length. Compared to using an irregularly spaced observation with  $M = 36$ , the RMS errors here are significantly lower, a difference of about 3 with both  $r^o = 6.9$  and of about 2.5 with  $r^o = 17.3$ , while the pattern correlations are much higher, showing an improvement of more than 10% in both cases of noise variance.

In the remainder of this section, we examine the numerical results in more detail. In Figure 2.16, we show the RMS errors of the unfiltered estimates, with linear interpolation on the top panel and kriging on the bottom panel, as functions of time for observation time interval  $T_{\text{obs}} = 0.01$ ,  $M = N^2 = 36$ , and  $r^o = 17.3$ . Notice that the RMS errors for both schemes oscillate and exhibit peaks at roughly the same times; maximum and minimum RMS errors tend to occur at the same times for both methods of interpolation.

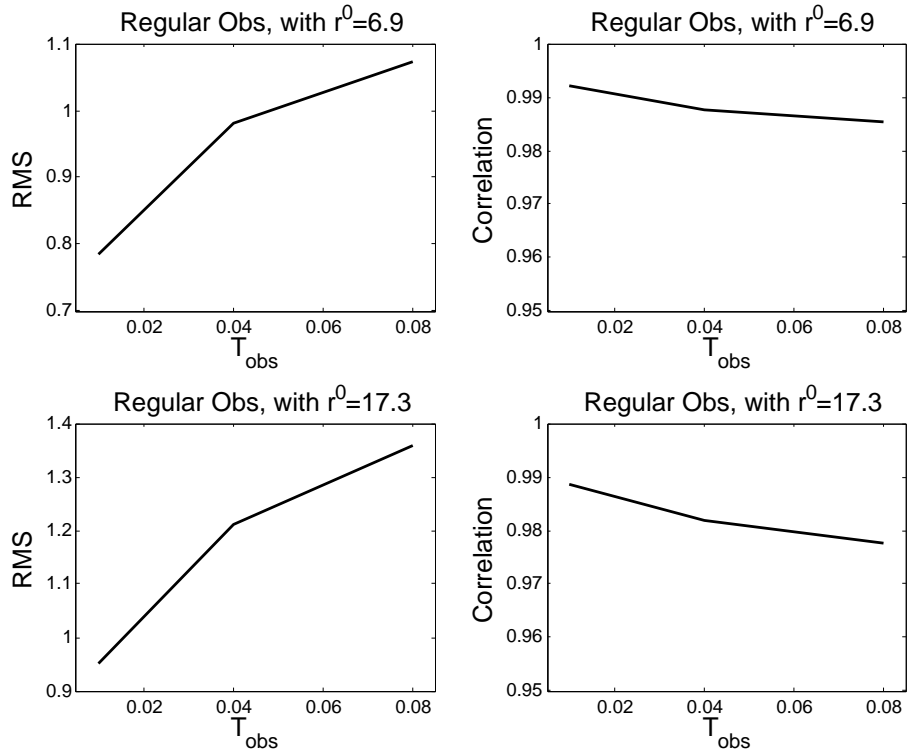


Figure 2.15: The average RMS errors (left panels) and pattern correlations (right panels) for numerical experiments with regularly spaced observations at the  $N \times N = 36$  model grid points, with noise variance  $r^o = 10\%E$  (top panels) and  $r^o = 25\%E$  (bottom panels).

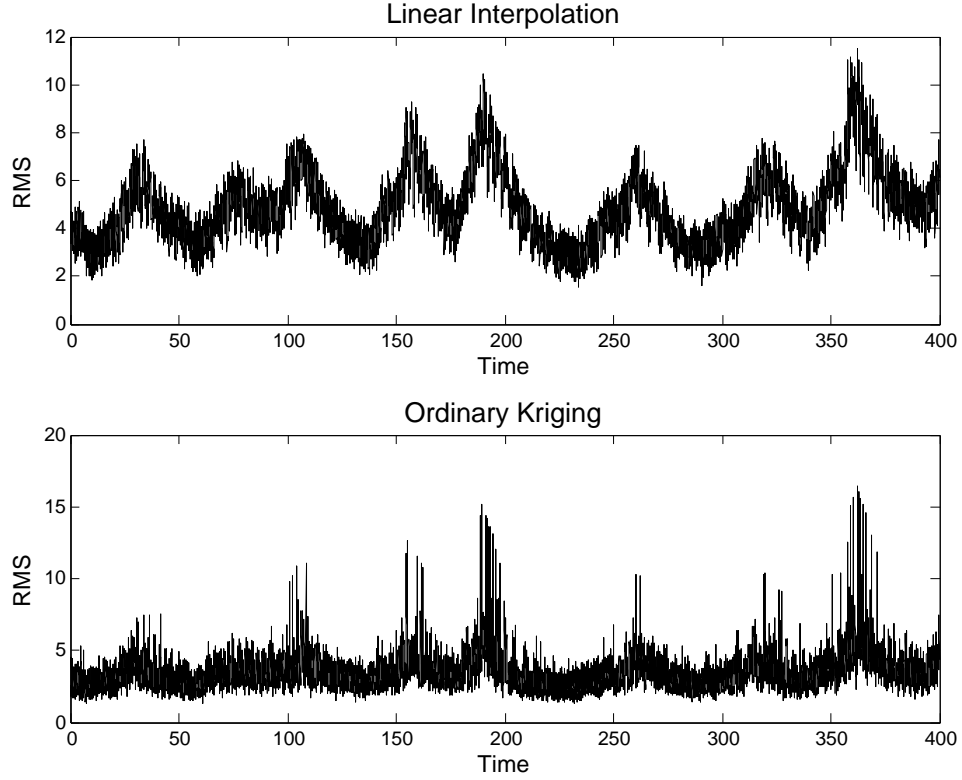


Figure 2.16: The RMS errors as functions of time, with  $M = N^2 = 36$ ,  $r^o = 17.3$  and  $T_{obs} = 0.01$ . The top panel shows the RMS error of the unfiltered linearly interpolated estimates and the bottom panel shows the RMS error of unfiltered kriging estimates.

To understand why the errors oscillate, we show the true streamfunction as well as the filtered and unfiltered streamfunction estimates of both interpolation schemes at the time  $T = 363.28$ , when the RMS error is at a maximum, in Figure 2.17. We also plot the truth and the estimates at time  $T = 61$ , when the RMS error is at a minimum in 2.18. Notice that there is a zonal jet in both instances, but more importantly, the contour scaling of  $\psi$  varies between  $\pm 6$  at time  $T = 61$  and between  $\pm 20$  at time  $T = 363.28$ . In either interpolation scheme, there is difficulty in capturing these peaks. However, the filtered estimate at time  $T = 363.28$  is more accurate compared to the filtered estimate at time  $T = 61$ , suggesting that the filter estimates are quite accurate when the signal-to-noise ration (between the true signal and the observation noises) are large. On the other hand, when the signal-to-noise ratio is small, the estimates are less accurate. In this case, one needs less noisy observations to improve the estimates.



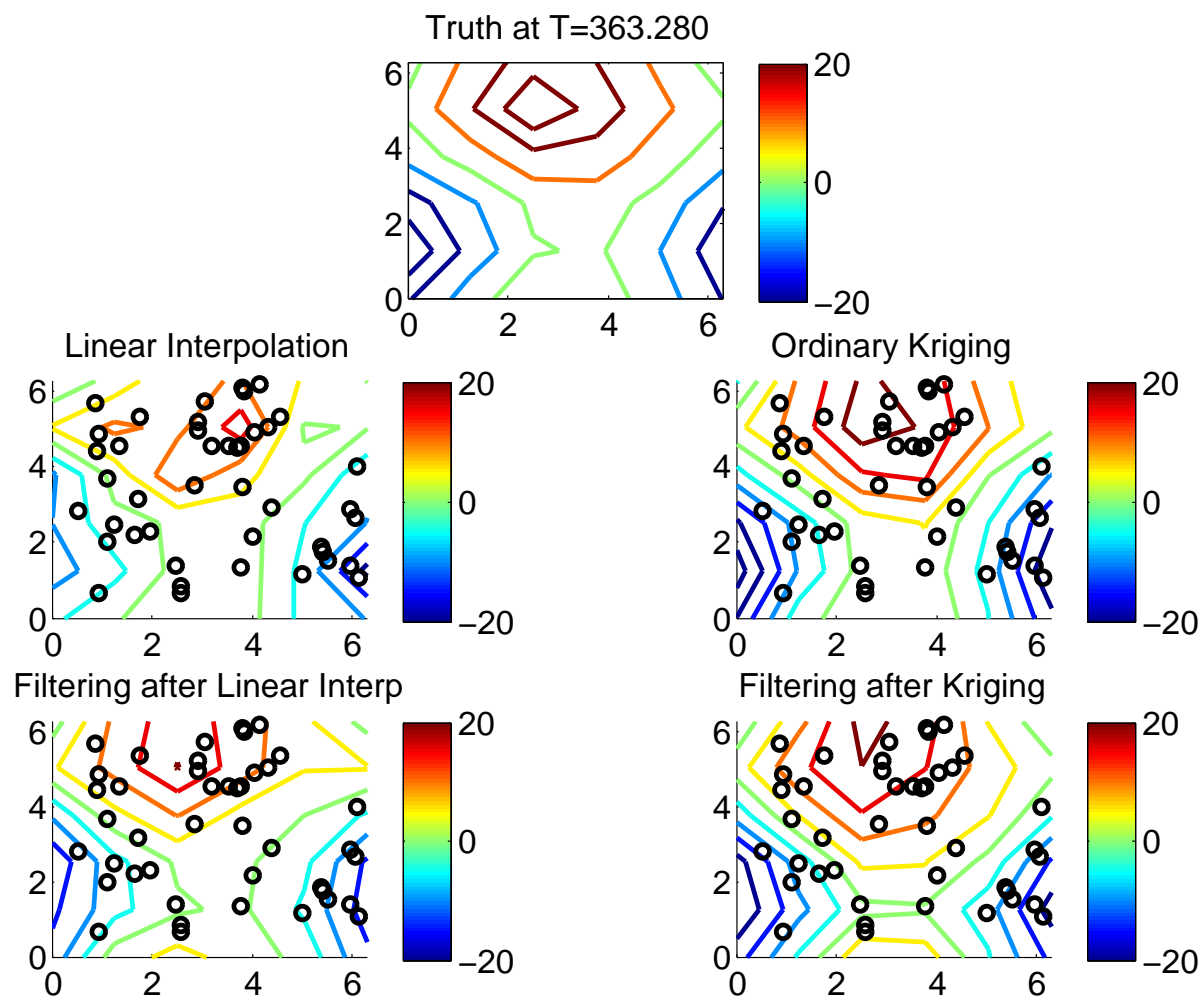


Figure 2.17: The truth (top panel) interpolated (middle panels) and filtered (bottom panels) fields at time  $T = 363.28$  (in contours). The circles depict observation locations.

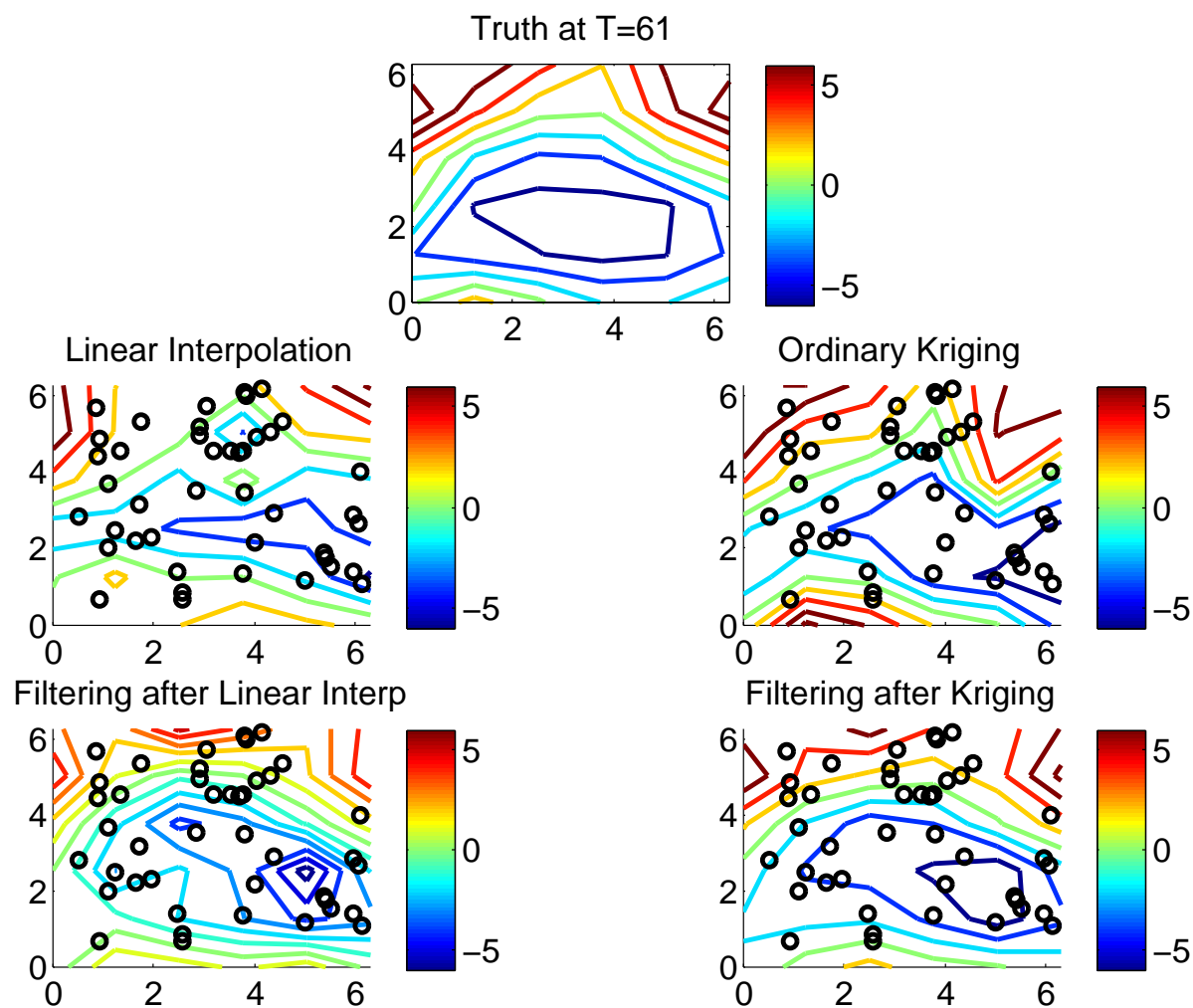


Figure 2.18: The truth, (top panel), interpolated (middle panels) and filtered (bottom panels) fields at time  $T = 61$  (in contours). The circles depict observation locations.

## 2.4 Summary

We investigated a hierarchical Bayesian approach for filtering irregularly spaced, sparse observations of geophysical turbulence. In particular, we blended an interpolation scheme, through either ordinary kriging or a deterministic piecewise linear interpolation, with a reduced stochastic filtering approach, the Fourier domain Mean Stochastic Model filter. The two-step hierarchical Bayesian approach proposed here interpolates raw data at irregularly spaced locations to regularly spaced filter model grid points and then assimilates this processed data set with the basic Kalman filter scheme on a diagonal Fourier domain mean stochastic model. We find that the dynamical constraint through MSM becomes more important when the observation noise variance is large, or when the observations are sparse. Second, when the observation network is very sparse, we find that both the linear interpolation and kriging schemes produce comparable posterior state estimates. Third, the two-step hierarchical Bayesian approach with kriging produces significantly improved filtered solutions when the observation network is spatially denser.

The filtered estimates with irregularly spaced observation networks are worse than the estimates with a regularly spaced observation network (exactly at the model  $N \times N = 36$  grid points). As we have mentioned, we need denser observations to improve the interpolated estimates in the first step of the proposed hierarchical Bayesian framework. For example, in the case of  $T_{obs} = 0.01$  and  $r^o = 17.3$ , the RMS error for the filtered kriging estimates with the irregularly spaced observations is about 3.6, much worse than the RMS error associated with the regularly spaced observations, 0.96. On the other hand, we should also mention that this case was considered in [25] with regularly spaced sparse observations at the filter model grid points. There, the authors reported filtered estimates with much higher average RMS error obtained from a perfect model simulation with the Local Least Squares-Ensemble Adjustment Kalman Filter, 6.76 (see Table 3 of [25]). Comparing these numerical results, we conclude that the hierarchical Bayesian approach is much better compared to the standard ensemble filtering approach that filters directly at observation locations. However, we anticipate that improved results with this framework are possible.

In our simulations, we assumed Gaussianity of the spatial distribution of the data, which may not be true in various applications. In these cases, one may want to consider statistical interpolations with appropriate distributions. For geostatistical data, however, kriging seems to be successful in practice [10]. Additionally, higher order deterministic interpolation schemes could easily introduce large correlations between different interpolated locations. Our two dimensional linear interpolation produces an interpolated error covariance matrix that is nearly diagonal in Fourier space and the MSM filter we used ignores these non-diagonal covariance terms. In contrast, kriging, by design, produces a Fourier domain diagonal interpolated error

covariance matrix. This is because kriging assumes stationarity and uses isotropic covariance models. As a consequence, the second step in our proposed hierarchical Bayesian approach can be performed without any additional approximation when the Fourier domain diagonal reduced stochastic models, such as MSM, are used with kriging. Finally, we expect that the results we obtained with linear interpolation are not robust on two-dimensional domains, because there are numerous criteria for constructing Lagrange functions. Similarly, the kriging results in this paper are based on a specific implementation of ordinary kriging with parametric covariance function of an exponential family. We suspect that improved estimates are possible with different classes, or different types of kriging [10].

## Chapter 3

# State Estimation from Cloudy Satellite Observations

The Atmospheric Infrared Sounder (AIRS) is a remote sensing instrument on NASA's Aqua satellite. This passive sensing instrument measures the intensity of infrared radiation of varying wavelengths at the top of the atmosphere, which depends on the chemical composition of the atmosphere within, as well as the atmospheric state, including its temperature and humidity. When it was launched in 2002, AIRS became the first of a new generation of sounders designed to provide higher resolution data, leading to better numerical weather prediction [38, 39]. The assimilation of AIRS observations that are unaffected by clouds has led to an improvement of weather forecast skill in both hemispheres. Compared to the High-Resolution Infrared Sounder (HIRS), currently operating on NOAA satellites, AIRS provides improved temperature accuracy, spatial resolution, and vertical resolution of atmospheric temperature profiles [38]. Furthermore, the assimilation of AIRS data alone produces better forecasts compared to single instrument experiments with HIRS data or Advanced Microwave Sounding Unit-A (AMSU-A) data, a microwave radiometer used on satellites [54].

The presence of a cloud can have a dramatic, undesired effect on passive satellite observations, as it can reflect or absorb electromagnetic radiation. In spite of this, it is not practical to use only cloud-free satellite observations; in fact less than 1% of AIRS observations have a clear field of view [13]. One strategy to improve this percent employs a process called cloud clearing [13, 15, 40]. This method uses neighboring cloud-free channels to estimate the radiance beneath the cloud. However, cloud clearing is not widely used, since the estimates tend to be biased towards cloud-free radiances, [52], and since the corresponding error covariance matrix is complicated [57]. Operationally, cloud-impacted AIRS data are not currently used in numerical weather prediction [38, 54, 60, 27]. Satellite observations are first subject to a cloud detection algorithm (see [20, 53] for descriptions of different approaches). If the field of view is found

to be cloudy, only AIRS channels which are determined to be insensitive to the temperature and humidity profiles below the cloud are used [27]; such channels do not suffer from the undesired effects of clouds, but tend to be uninformative. Unfortunately, the percentage of AIRS measurements that are used in numerical weather prediction remains small.

In this chapter, we will experiment with the assimilation of synthetic cloudy AIRS data. Recently, there has been evidence to suggest that data assimilation of cloudy radiances together with clear sky radiances does have a positive impact on forecast skill [52, 56, 59]. Cloud-impacted radiances contain considerable data that could enhance weather prediction. Our goal is to assess the potential of using cloudy AIRS data in numerical weather prediction and to study the biases that arise due to the presence of a cloud. In particular, we will examine the sensitivity of the analysis skill with respect to cloud top pressure and cloud cover percent. In an ideal scenario, we want to know the best filter analysis that we can obtain from cloudy data. To gauge the effectiveness of the analysis step, we include the results of several numerical experiments conducted solely with simulated clear sky AIRS observations, and solely with direct observations of temperature and humidity.

To generate synthetic AIRS observations, we will use a numerical simulation of the multi-cloud model, a toy model of tropical convection introduced in [32, 33], to specify the true atmospheric state. Modeling tropical weather is complex because of the variety of interacting nonlinear processes over numerous spatio-temporal scales [34]. This toy model captures the dynamics of a three cloud convective structure that occurs in the tropics [41, 29, 67, 63, 21, 12, 35, 69]: low lying congestus clouds which moisten the lower troposphere, ensuing penetrative cumulus clouds associated with deep convection, and stratiform clouds which remain afterward. There are three multcloud variables, congestus heating, stratiform heating, and precipitation rate, which correlate with the type of cloud present at each time and location and its coverage percent. Additionally, this model includes a switching function which turns off deep convection and amplifies congestus heating when the atmosphere is too dry, and promotes deep convective heating when the atmosphere is moist. We use the three heating rates along with the switch function to simulate the effects of clouds in AIRS measurements.

We use the prototype community Radiative Transfer Model (pCRTM), [22], to produce synthetic AIRS observations of brightness temperature at the top of the atmosphere in the presence of clouds. The pCRTM uses atmospheric state values of temperature and humidity in discrete vertical layers of the atmosphere to produce estimates of radiance intensities associated with 281 different wavelengths in the infrared spectrum. Subsequently, it converts these radiances to brightness temperatures. We modify the pCRTM to include a cloud cover percent in each atmospheric layer that lies beneath a cloud (if there is no cloud present, we use a cloud cover percent of zero). Using the local ensemble transform Kalman filter (LETKF), [6, 28], we assimilate these cloudy observations with the multcloud model, extending the work in assimilating

clear sky radiances with the LETKF [17, 4].

In Section 3.1, we review the dynamical multcloud model that mimics the moist convective tropical dynamics. In Section 3.2, we review the radiative transfer model that generates brightness temperatures based on the solutions of the multcloud model. Subsequently, in section 3.3, we describe how we generate synthetic AIRS data in the clear sky and cloudy sky configurations. In Section 3.4, we describe the local ensemble transform Kalman filter and the the biases that can occur when assimilating cloudy data. In Section 3.5, we present our numerical results. We close with a summary and discussion in In Section 3.6.

## 3.1 The Multicloud Model

In our numerical experiments, we use a simulation of the multcloud model as the truth, and as the forecast model. Therefore, these are perfect model experiments [9], because the forecast model is the same model that governs the true system dynamics. There are three cloud types that are prevalent in the tropical boundary layer, shallow congestus clouds, stratiform clouds, and deep cumulus clouds. Congestus clouds by themselves are characteristic of unstable atmospheric regions undergoing convection or heating from the moisture in the atmosphere. These clouds moisten the lower troposphere and give rise to the deep convective clouds. Stratiform precipitation, in turn, dries the atmosphere, bringing this cycle to an end. Stratiform clouds, which have a uniform base but great height, remain. The multcloud model captures these cloud dynamics through the parametrization of two convectively coupled heating modes: a deep-convective mode and a stratiform mode. This model is unique because it includes cumulus congestus clouds and factors in the amount of moisture in the middle of the troposphere with a switch parameter  $\Lambda$ , in order to trigger deep convection or shut it off appropriately [32]; if the middle troposphere is dry, congestus heating is forced to moisten the lower troposphere. Conversely, if the middle troposphere is moist, deep convection is amplified and congestus convection is shut off.

### 3.1.1 Background Information

When a fluid parcel changes pressure adiabatically (without exchanging heat with its surroundings), its temperature will change, but not because of heating. The potential temperature  $\theta$  is defined as the temperature that a fluid parcel would have if it moved adiabatically to some reference pressure (typically 1,000 mb, the pressure at Earth’s surface) [65]. In adiabatic flow, potential temperature is conserved, that is, for a flow with velocity  $\vec{u}$ ,

$$\frac{D\theta}{Dt} = \frac{\partial\theta}{\partial t} + \vec{u} \cdot \nabla\theta = 0.$$

The multicloud model incorporates potential temperature, as well as equivalent potential temperature,  $\theta_e$ , the potential temperature that a fluid parcel achieves if its moisture condenses and all of the heat released goes into heating the parcel. Equivalent potential temperature is defined so that it is approximately conserved during condensation [65].

The hydrostatic Boussinesq primitive equations are the basis for the multicloud model. The primitive equations, which approximate atmospheric flow, consist of three parts: conservation of momentum, conservation of mass, and a thermal energy equation to close the system. There are several approximations made in the primitive equations: Earth's curvature is approximated by using a domain reduced to a periodic strip along the equator; the vertical direction  $z$  is much smaller than the radius of the Earth (the shallow fluid approximation); density variations are negligible, except in the buoyancy term. Using these approximations, the primitive equations are given in [34] by

$$\begin{aligned}\frac{\partial \vec{v}}{\partial t} + \vec{v} \cdot \nabla \vec{v} + w \frac{\partial \vec{v}}{\partial z} + \beta y \vec{v}^\perp &= -\nabla p + S_{\vec{v}}, \\ \frac{\partial p}{\partial z} &= \frac{g\Theta}{\theta_o}, \\ \frac{\partial \Theta}{\partial t} + \vec{v} \cdot \nabla \Theta + w \frac{\partial \Theta}{\partial z} + \frac{N^2 \theta_o}{g} w &= S_\Theta, \\ \nabla \cdot \vec{v} + \frac{\partial w}{\partial z} &= 0.\end{aligned}$$

Here  $\vec{v} = (u, v)$  is the horizontal velocity field, where  $u$  is the zonal velocity (east/west) and  $v$  is the meridional velocity (north/south);  $w$  is the vertical velocity; scalars  $\Theta$  and  $p$  are potential temperature and pressure perturbations;  $N$  is the Brunt-Vaisala buoyancy frequency, the frequency with which a vertically displaced parcel would oscillate around its equilibrium position under the influence of buoyancy and no other forces;  $\vec{v}^\perp = (-v, u)$ ;  $\beta = 2.28 \times 10^{-11} m^{-1} s^{-1}$  is the gradient of the Coriolis parameter at the equator;  $S_{\vec{v}}$  and  $S_\Theta$  are the sources and sinks of momentum and potential temperatures. The primitive equations are supplemented with a rigid lid boundary condition, in which vertical velocity is zero at the surface and at the top of the troposphere.

The Coriolis parameter quantifies the effects of Earth's rotation as  $f = 2\Omega \sin \phi$ . At the equator,  $\phi = 0$  and the Coriolis force vanishes, meaning that the Earth's rotation does not affect atmospheric and oceanic dynamics. This gives rise to several types of waves with complex meridional and zonal structure. We distinguish barotropic waves, which are capable of propagating poleward, from baroclinic waves, which can only propagate along the equator [34]. Barotropic waves have no dependence on height  $z$ , while baroclinic waves do.

We can decompose the primitive equations into barotropic and baroclinic parts. Because of the rigid lid boundary condition, vertical velocity and potential temperature terms are expanded



in terms of  $\{\sin(m\pi z/H_T)\}_{m=0}^\infty$ , where  $H_T \approx 16$  km is the height of the tropical troposphere. Horizontal velocity and pressure terms are expanded in terms of  $\{\cos(m\pi z/H_T)\}_{m=0}^\infty$ . Since barotropic flow has no height dependence,  $m = 0$  corresponds to the barotropic mode. Let variables with an overhead bar denote barotropic modes. Then the truncated primitive equations projected onto the barotropic and first two baroclinic modes are given by

$$\begin{aligned} \begin{pmatrix} V \\ p \end{pmatrix} (x, y, z, t) &= \begin{pmatrix} \bar{v} \\ \bar{p} \end{pmatrix} (x, y, t) \\ &+ \begin{pmatrix} \vec{v}_1 \\ p_1 \end{pmatrix} (x, y, t) \sqrt{2} \cos\left(\frac{\pi z}{H_T}\right) + \begin{pmatrix} \vec{v}_2 \\ p_2 \end{pmatrix} (x, y, t) \sqrt{2} \cos\left(\frac{2\pi z}{H_T}\right) \end{aligned} \quad (3.1)$$

and

$$\begin{pmatrix} W \\ \Theta \end{pmatrix} (x, y, z, t) = \begin{pmatrix} w_1 \\ \theta_1 \end{pmatrix} (x, y, t) \sqrt{2} \sin\left(\frac{\pi z}{H_T}\right) + \begin{pmatrix} w_2 \\ \theta_2 \end{pmatrix} (x, y, t) \sqrt{2} \sin\left(\frac{2\pi z}{H_T}\right). \quad (3.2)$$

### 3.1.2 The Model

The multicloud model is a Galerkin projection of the primitive equations onto the first two baroclinic modes under consideration. This means that we use the primitive equations to describe the first two baroclinic terms in (3.1)-(3.2). We ignore the barotropic term, so that we only consider equatorially trapped waves in the solution. The first mode directly heats the system through precipitation from the deep penetrative clouds. The second baroclinic mode heats the stratiform clouds while cooling the congestus clouds, and vice versa.

Let  $\vec{v}_j = (u_j, v_j)$  for  $j = 1, 2$  denote the first and second baroclinic mode wind velocity and let  $\theta_j$  for  $j = 1, 2$  denote the potential temperature associated with each mode. With the Galerkin projection, interactions between the two modes becomes a coupled system of shallow water equations corresponding to a convective heating mode and a stratiform and congestus mode:

$$\begin{aligned} \frac{\partial \vec{v}_j}{\partial t} + \beta y \vec{v}_j^\perp - \nabla \theta_j &= -c_d(u_o) \vec{v}_j - \frac{1}{\tau_R} \vec{v}_j \\ \frac{\partial \theta_1}{\partial t} - \text{div}(\vec{v}_1) &= \frac{\pi}{2\sqrt{2}} P + S_1 \\ \frac{\partial \theta_2}{\partial t} - \frac{1}{4} \text{div}(\vec{v}_2) &= \frac{\pi}{2\sqrt{2}} (H_c - H_s) + S_2, \end{aligned}$$

where  $P \geq 0$  gives the heating rate from deep convection,  $H_c$  and  $H_s$  are the congestus and stratiform heating rates, respectively,  $c_d$  is the boundary layer turbulence momentum friction,  $u_o = 2$  m/s is the turbulent velocity scale,  $S_j$  is a radiative cooling rate associated with baroclinic

mode  $j$ , and  $\vec{v}_j^\perp = (-v_j, u_j)$ . The nonlinearities of this model lie in the heating rates. The details of this derivation can be found in [55, 18, 49]. Notice that the congestus clouds heat the second baroclinic mode at rate  $H_c$  from below and that the stratiform clouds heat this mode from above at rate  $H_s$ . This mode is therefore cooled by  $H_c$  from above and by  $H_s$  from below. We ignore the meridional features of the tropical disturbances, thus eliminating the Coriolis term and variations in  $y$ . This allows us to simplify these equations to

$$\begin{aligned}\frac{\partial u_j}{\partial t} - \frac{\partial \theta_j}{\partial x} &= -c_d(u_o)u_j - \frac{1}{\tau_R}u_j \\ \frac{\partial \theta_1}{\partial t} - \frac{\partial u_1}{\partial x} &= \frac{\pi}{2\sqrt{2}}P + S_1 \\ \frac{\partial \theta_2}{\partial t} - \frac{1}{4}\frac{\partial u_2}{\partial x} &= \frac{\pi}{2\sqrt{2}}(H_c - H_s) + S_2.\end{aligned}\tag{3.3}$$

The radiative cooling rates  $S_1$  and  $S_2$  are given by the Newtonian cooling model

$$S_j = -Q_{R,j}^o - \frac{1}{\tau_D}\theta_j, \quad j = 1, 2,\tag{3.4}$$

where  $Q_{R,j}^o$  are the radiative cooling rates at radiative convective equilibrium.

The equivalent potential temperature at the boundary layer,  $\theta_{eb}$ , and the vertically integrated moisture content (humidity),  $q$ , are included in this model and are respectively given by

$$\begin{aligned}\frac{\partial \theta_{eb}}{\partial t} &= \frac{1}{h_b}(E - D) \\ \frac{\partial q}{\partial t} + \frac{\partial}{\partial x}[(u_1 - \tilde{\delta}u_2)q] + \tilde{Q}\left(\frac{\partial u_1}{\partial x} + \tilde{\lambda}\frac{\partial u_2}{\partial x}\right) &= \frac{D}{H_T} - P.\end{aligned}\tag{3.5}$$

A detailed derivation of the moisture equation is given in [31]. In these equations,  $h_b \approx 500\text{m}$  is the height of the moisture boundary layer, and  $\tilde{Q}$ ,  $\tilde{\lambda}$  and  $\tilde{\delta}$  are parameters of the moisture background and perturbation vertical profiles. Additionally,  $E$  represents the sea surface evaporative heating and  $D$  represents downdrafts, both of which influence  $\theta_{eb}$ . This model differs from others in its use of deep convective heating/precipitation  $P$  and downdrafts  $D$ , and because it includes congestus heating  $H_c$  the in equation for  $\theta_2$ .

The surface evaporation  $E$  obeys the relation

$$\frac{1}{h_b}E = \frac{1}{\tau_e}(\theta_{eb}^* - \theta_{eb}),\tag{3.6}$$

where  $\tau_e$  is the evaporative time scale and  $\theta_{eb}^*$  is the boundary layer saturation equivalent potential temperature. Notice in this equation that evaporation will occur until the boundary

layer is saturated, when  $\theta_{eb} = \theta_{eb}^*$ . The value of  $\theta_{eb}^*$  on a warm ocean surface is fixed such that at radiative convective equilibrium  $\theta_{eb}^* - \bar{\theta}_{eb} = 10$  K, [32], where  $\bar{\theta}_{eb}$  is the equilibrium value of  $\theta_{eb}$ . Over our domain, which lies along the equator and has length 40,000 km, the value of  $\theta_{eb}^*$  is raised from -5 K to 5 K in the center of a 10,000 km warm pool region, centered at 20,000 km.

In the mid troposphere, the equivalent potential temperature anomaly,  $\theta_{em}$ , can be approximated by

$$\theta_{em} \approx q + \frac{2\sqrt{2}}{\pi}(\theta_1 + \alpha_2\theta_2), \quad (3.7)$$

where  $\alpha_2 = 0.1$  allows for a small contribution of the second baroclinic mode temperature to  $\theta_{em}$ .

To quantify the level of moisture or dryness in the middle troposphere, the multicloud model also uses a switch parameter  $\Lambda$  [32]. When the equivalent potential temperature at the boundary layer and at the middle of the troposphere differ by more than a certain threshold  $\theta^+$ , the atmosphere is dry and we set  $\Lambda = 1$ . When this difference is below a lower threshold  $\theta^-$ , the atmosphere is wet and we set  $\Lambda = \Lambda^* < 1$ . We will take  $\Lambda^* = 0.2$ ,  $\theta^+ = 20$  K and  $\theta^- = 10$  K, as in [32]. These values come from climatological values recorded in the tropics. Between  $\theta^+$  and  $\theta^-$ ,  $\Lambda$  is linearly interpolated so that

$$\Lambda = \begin{cases} 1 & \text{if } \theta_{eb} - \theta_{em} > \theta^+ \\ 0.08(\theta_{eb} - \theta_{em}) - 0.6 & \text{if } \theta^- \leq \theta_{eb} - \theta_{em} \leq \theta^+ \\ 0.2 & \text{if } \theta_{eb} - \theta_{em} < \theta^-. \end{cases} \quad (3.8)$$

This equation for  $\Lambda$  only holds for our values of  $\Lambda^*$ ,  $\theta^+$ , and  $\theta^-$ . With this switch function we can write precipitation and downdrafts as

$$P = \frac{1 - \Lambda}{1 - \Lambda^*} P_o, \quad \text{and} \quad D = \Lambda D_o,$$

while the stratiform and congestus heating rates  $H_s$  and  $H_c$  satisfy

$$\frac{\partial H_s}{\partial t} = \frac{1}{\tau_s}(\alpha_s P - H_s) \quad (3.9)$$

and

$$\frac{\partial H_c}{\partial t} = \frac{1}{\tau_c} \left( \alpha_c \frac{\Lambda - \Lambda^*}{1 - \Lambda^*} \frac{D}{H_T} - H_c \right). \quad (3.10)$$

If the middle troposphere is dry,  $\Lambda = 1$  and  $P = 0$ , so that (3.9) reduces to

$$\frac{\partial H_s}{\partial t} = -\frac{H_s}{\tau_s}$$

and (3.10) reduces to

$$\frac{\partial H_c}{\partial t} = \frac{1}{\tau_c} \left( \frac{\alpha_c D_o}{H_T} - H_c \right).$$

The stratiform heating  $H_s$  decreases while the congestus heating  $H_c$  increases, for a time. While the middle troposphere is dry, downdraft  $D$  is at a maximum and  $P = 0$  so that  $\partial q / \partial t$  in (3.5) is positive, increasing the moisture content; the shallow clouds moisten and precondition the middle troposphere to sustain deep convection and promote boundary layer clouds. This in turn will lower the difference in potential temperatures at the boundary layer and middle troposphere,  $\theta_{eb} - \theta_{em}$ , below the threshold  $\theta^+$  and thus the switch function will move towards its lower value  $\Lambda^*$ . When  $\Lambda = \Lambda^*$ , precipitation  $P = P_o$  is at a maximum while downdraft  $D$  is at a minimum. This time stratiform heating is favored and by (3.5) the moisture content decreases. This leads to an increase in the difference between the boundary layer and middle troposphere temperature, increasing  $\Lambda$  again.

The quantity  $P_o$  gives the maximum possible value of deep convective precipitation  $P$  and can be given by

$$P_o = \frac{1}{\tau_{\text{conv}}} [a_1 \theta_{eb} + a_2 (q - \hat{q}) - a_o (\theta_1 + \gamma_2 \theta_2)]^+ \quad (3.11)$$

where  $\hat{q}$  is a threshold constant related to the level of saturation in the troposphere,  $\tau_{\text{conv}}$  is the convective time scale and  $f^+ = \max\{f, 0\}$ . The coefficient  $a_o$ , is a parameter associated with the inverse buoyancy relaxation time in [19].

The quantity  $D_o$  represents the maximum possible downdrafts. The downdrafts are closed by

$$D_o = \frac{m_o}{\bar{P}} [\bar{P} + \mu_2 (H_s - H_c)]^+ (\theta_{eb} - \theta_{em}), \quad (3.12)$$

where  $m_o$  scales the downdraft mass flux,  $\mu_2$  is a parameter that permits the formation of stratiform and congestus mass flux anomalies [44], and  $\bar{P}$  is the precipitation heating rate at equilibrium. For easy reference, we give a table of parameters and their values in Table 3.1 [32].

We filter eight multcloud prognostic variables,  $u_j$  and  $\theta_j$ , for  $j = 1, 2$ ,  $\theta_{eb}$ ,  $q$ ,  $H_c$ , and  $H_s$ , resolved at 1,000 discrete grid points on a one-dimensional periodic domain in the equator, with a spatial resolution of 40 km. These variables have nondimensional units. The length scale is defined by the equatorial Rossby deformation radius (the scale at which Earth's rotation becomes significant to the system dynamics) as  $L_d \approx 1500$  km, while time scale is given by  $T = L_d / c \approx 8$  hours. The first baroclinic dry gravity wave speed  $c \approx \text{m/s}$ , gives the velocity scale, and the dry-static stratification  $\bar{\alpha} = (H_T N^2 \theta_o / \pi g) \approx 15$  K is the temperature scale.

This model reproduces a number of observational features of equatorial convectively coupled waves. With the parameters in [46], the model reproduces the Madden-Julian Oscillation (MJO). The MJO is an equatorial, planetary-scale wave envelope of complex multiscale convective processes that begins as a standing wave in the Indian Ocean and propagates east across the

Table 3.1: Parameters and values of the multicloud model

Parameter	Value	Description
$H_T$	16 km	Height of tropical troposphere
$Q$	0.9	Moisture stratification factor
$\tilde{\lambda}$	0.8	2 <sup>nd</sup> baroclinic contribution to moist. conv. from background
$\tilde{\delta}$	0.1	2 <sup>nd</sup> baroclinic contribution to moist. conv. from anomalies
$\tau_R$	50 days	Rayleigh wind friction relaxation time
$\tau_D$	75 days	Newtonian cooling relaxation time
$c_d$	0.001	Boundary layer turbulence momentum friction
$L_e$	1500 km	Equatorial deformation radius, length scale
$c$	50 m/s	Speed of 1 <sup>st</sup> baroclinic gravity wave, velocity scale
$T = L_e/c$	8 hours	Time scale
$\bar{\alpha}$	15 K	Dry static stratification, temperature scale
$N$	0.01 s <sup>-1</sup>	Brunt-Vaisala buoyancy frequency
$\theta_o$	300 K	Reference temperature
$h_b$	500 m	Boundary layer height
$\alpha_2$	0.1	Contribution of $\theta_2$ to middle troposphere $\theta_e$
$\theta_{eb}^*$	10 K + $\theta_{eb}$	Boundary layer equivalent potential temperature
$\tau_e$	8 h	Evaporative time scale in boundary layer
$\theta^\pm$	10, 20 K	Temp thresholds used to define $\Lambda$
$\Lambda^*$	0.2	Lower threshold of $\Lambda$
$\tau_s$	3 hours	Stratiform heating adjustment time
$\alpha_s$	0.25	Stratiform heating adjustment heating coefficient
$\tau_c$	1 hour	Congestus heating adjustment time
$\alpha_c$	0.5	Congestus heating adjustment coefficient
$a_o$	7.5	Inverse buoyancy time scale
$a_1$	0.1	Contribution of $\theta_{eb}$ to convective parametrization
$a_2$	0.9	Contribution of $q$ to convective parametrization
$\tau_{\text{conv}}$	2 hours	Deep convective reference time scale
$\gamma_2$	0.1	Strength of lower troposphere coupling
$\bar{\theta}_{eb} - \bar{\theta}_{em}$	14 K	Discrepancy at RCE
$\mu_2$	0.5	Contribution of stratiform/congestus mass flux anomalies

Pacific at roughly 5 m/s [69]. The MJO remains unexplained, but important to tropic dynamics, as it is responsible for intraseasonal variability in the tropics [69]. In [1], they simulated a realistic MJO, using a coarse resolution aquaplanet global circulation model, coupled with a multicloud model parametrization.

We give an example simulation with bulk parameters  $\tilde{Q} = 0.9$ ,  $\tau_R = 50$  days,  $\bar{\theta}_{eb} - \bar{\theta}_{em} = 14$  K and the convective parameters in Table 3.1, as in [26]. The intraseasonal timescale is generated

using  $\tau_{\text{conv}} = 2$  hours and  $\tau_s = 3$  hours and  $\tau_c = 1$  hour. In Figure 3.1, plot the precipitation contours as a function of space and time. We observe both eastward and westward propagating waves, traveling with speed 6.1 m/s. At the center of the warm pool, at 20,000 km, there are regions of strong convective activity.

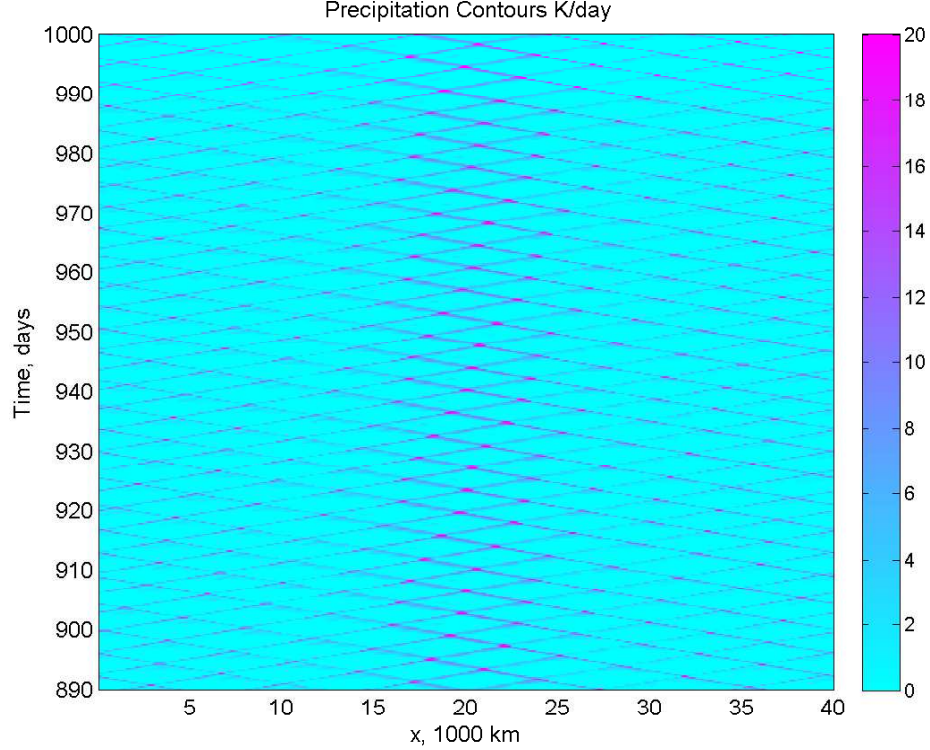


Figure 3.1: Contour plot of deep convective heating  $P$  from a numerical simulation of the multicloud model.

## 3.2 Radiative Transfer Model

The Radiative Transfer Model gives the intensity of electromagnetic radiation observed at the top of Earth's atmosphere. We use this model to generate synthetic satellite observations.

In a non-scattering atmosphere in local thermal equilibrium, [42] gives the basic equation

for observed intensity of wavelength  $\lambda$ ,  $I_\lambda$ , as

$$\mu \frac{dI_\lambda(\tau, \mu)}{d\tau} = I_\lambda(\tau, \mu) - B_\lambda(T(\tau)), \quad (3.13)$$

where  $\mu = \cos \theta$ ,  $\theta$  is the emergent angle,  $\tau$  is optical depth, and  $B_\lambda(T)$  is Planck's function. Planck's function is a relation describing the amount of electromagnetic radiation of a particular wavelength emitted from a black body, a hypothetical object which can absorb all incident radiation, regardless of wavelength. If a black body of temperature  $T$ , in Kelvin, emits radiation of wavelength  $\lambda$ , the energy of this radiation is given by

$$B_\lambda(T) = \frac{2hc^2\lambda^{-5}}{e^{hc/k\lambda T} - 1}$$

with Planck constant  $h = 6.626 \times 10^{-34} Js$ , Boltzmann constant  $k = 1.3806 \times 10^{-23} J/deg$  and speed of light  $c = 3 \times 10^8$  m/s. Notice that if the intensity of radiation is measured, one can invert the Planck function to compute a temperature  $T$ . In this case,  $T$  is referred to as brightness temperature. The satellite observations we use are of brightness temperatures, calculated from radiance intensity measured at the top of the atmosphere. In the remainder of this chapter, we use the terms brightness temperatures and radiances interchangeably.

The optical depth  $\tau$  in (3.13) is a dimensionless measure of how much radiation is absorbed when passing through a portion of the atmosphere. Optical depth increases with path length, and depends on the density of atmospheric gases. Different atmospheric gases will absorb radiation of different wavelengths. For example, carbon dioxide will absorb radiation of wavelength equal to 4.2 micrometers and 15 micrometers. Optical depth is defined by

$$\tau = \int_z^{z_\infty} \kappa_\lambda(z') \rho_a(z') dz' \quad (3.14)$$

where  $\rho_a$  and  $\kappa_\lambda$  are the density and absorption coefficient of the absorbing gas, respectively.

The solution of (3.13) is

$$I_\lambda(\tau, \mu) = I_\lambda(\tau^*) e^{-(\tau^* - \tau)/\mu} + \int_\tau^{\tau^*} \mu^{-1} B_\lambda(T(\tau')) e^{-(\tau' - \tau)/\mu} d\tau', \quad (3.15)$$

in which

$$\tau^* = \int_0^{z_\infty} \kappa_\lambda(z') \rho_a(z') dz'$$

is the optical depth at the surface and  $I_\lambda(\tau^*)$  is the radiance emitted from the surface. Notice that temperature  $T$  is a function of optical depth. This is because the temperature profile of the atmosphere depends on optical depth. Let us consider only the radiation of upwelling direction,

so that  $\theta = 0$  and  $\mu = 1$ . Let  $\epsilon_\lambda$  be the surface emissivity, a measure of the Earth's surface ability to emit radiation of wavelength  $\lambda$  (for a perfect black body,  $\epsilon = 1$ ). Then we can assume that  $I_\lambda(\tau^*) = \epsilon_\lambda B_\lambda(T_s)$ , where  $T_s$  is the surface temperature, in Kelvin. AIRS takes measurements in the thermal IR region, with wavelengths between 3.5 and 20 micrometers, in which  $\epsilon_\lambda$  is close to 1 and  $I_\lambda(\tau^*) \approx B_\lambda(T_s)$ .

Define monochromatic transmittance  $T_\lambda$  as

$$T_\lambda(\tau) = e^{-\tau}, \quad \frac{dT_\lambda}{d\tau} = -e^{-\tau}.$$

Then we have

$$\begin{aligned} I_\lambda(\tau, 1) = I_\lambda(\tau) &= I_\lambda(\tau^*)T_\lambda(\tau^* - \tau) + \int_\tau^{\tau^*} B_\lambda(T(\tau'))T_\lambda(\tau' - \tau)d\tau' \\ &= B_\lambda(T_s)T_\lambda(\tau^* - \tau) + \int_{\tau^*}^\tau B_\lambda(T(\tau'))\frac{\partial T_\lambda(\tau' - \tau)}{\partial \tau'}d\tau'. \end{aligned}$$

At the top of the atmosphere,  $\tau = 0$  and (3.15) becomes

$$I_\lambda(0) = B_\lambda(T_s)T_\lambda(\tau^*) + \int_{\tau^*}^0 B_\lambda(T(\tau))\frac{\partial T_\lambda(\tau)}{\partial \tau}d\tau. \quad (3.16)$$

If we assume that the atmosphere is in hydrostatic balance, meaning that the pressure gradient balances with gravity as

$$\frac{\partial p}{\partial z} = -g\rho,$$

then we can use pressure to define the vertical coordinate axis. Using hydrostatic balance and the mixing ratio  $q = \rho_a/\rho$ , the density of the absorbing gas is given by

$$\rho_a = q\rho = -\frac{q}{g}\frac{\partial p}{\partial z}$$

and optical depth at the surface is given by

$$\tau^* = \int_0^{p_s} \kappa_\lambda(p) \frac{q(p)}{g} dp.$$

We also have

$$\begin{aligned} \frac{\partial T_\lambda(p)}{\partial p} &= \frac{\partial T_\lambda(p)}{\partial \tau} \frac{d\tau}{dp} \\ &= \frac{\partial T_\lambda}{\partial \tau} \left( \frac{\kappa_\lambda q}{g} \right) \end{aligned}$$



so that

$$\frac{\partial T_\lambda(\tau)}{\partial \tau} d\tau = \frac{\partial T_\lambda(p)}{\partial p} dp.$$

Thus (3.16) becomes

$$I_\lambda(0) = B_\lambda(T_s)T_\lambda(p_s) + \int_{p_s}^0 B_\lambda[T(p)] \frac{\partial T_\lambda(p)}{\partial p} dp \quad (3.17)$$

This equation describes the observed intensity of electromagnetic radiance of wavelength  $\lambda$  at the top of the atmosphere in a clear field of view. The function

$$\frac{\partial T_\lambda(p)}{\partial p} = \frac{\kappa_\lambda q}{g} \exp \left( \int_0^p \kappa_\lambda(p') \frac{q(p')}{g} dp' \right)$$

in the integrand is called the weighting function. It weights the Planck function depending on the density of the absorbing gas present at different layers in the atmosphere. Example weighting functions are shown in Figure 3.2 for three different AIRS channels. The AIRS instrument has 281 infrared channels, each of which detects radiation of a different wavelength. The first panel of Figure 3.2 shows the weighting function corresponding to channel frequency  $680.14 \text{ cm}^{-1}$ . In this case, the absorbing gases are dense at the top of the atmosphere, and therefore all observed radiance at the top of the atmosphere emanates from this layer of the atmosphere. The second panel shows the weighting function corresponding to channel frequency  $1377.43 \text{ cm}^{-1}$ . This weighting function peaks at the top of the atmosphere, and indicates that the absorbing gas is present in the upper and middle atmospheric layers. The third panel shows the weighting function corresponding to channel frequency  $2500.6 \text{ cm}^{-1}$ . The weighting function spreads out over the entire atmosphere, and tapers off close to the surface. For this wavelength, the absorbing gas is present throughout the atmosphere, but to a lesser slightly extent close to the surface.

### 3.2.1 Adding Cloud Cover

Consider a field of view consisting of a cloud layer located at pressure  $p_c$ . Denote the fraction of cloud cover as  $\eta$ , the temperature at the cloud top as  $T_c$ , the surface pressure  $p_s$ , and the cloud emissivity  $\epsilon_\lambda$ . Then the radiance observed by the satellite can be written as

$$\begin{aligned} I_\lambda^{cld} &= (1 - \eta\epsilon_\lambda) \left[ B_\lambda(T_s)T_\lambda(p_s, p_c) + \int_{p_s}^{p_c} B_\lambda[T(p)] \frac{\partial T_\lambda(p, p_c)}{\partial p} dp \right] T_\lambda(p_c, 0) \\ &+ \eta\epsilon_\lambda B_\lambda(T_c)T_\lambda(p_c, 0) + \int_{p_c}^0 B_\lambda[T(p)] \frac{\partial T_\lambda(p, 0)}{\partial p} dp. \end{aligned} \quad (3.18)$$

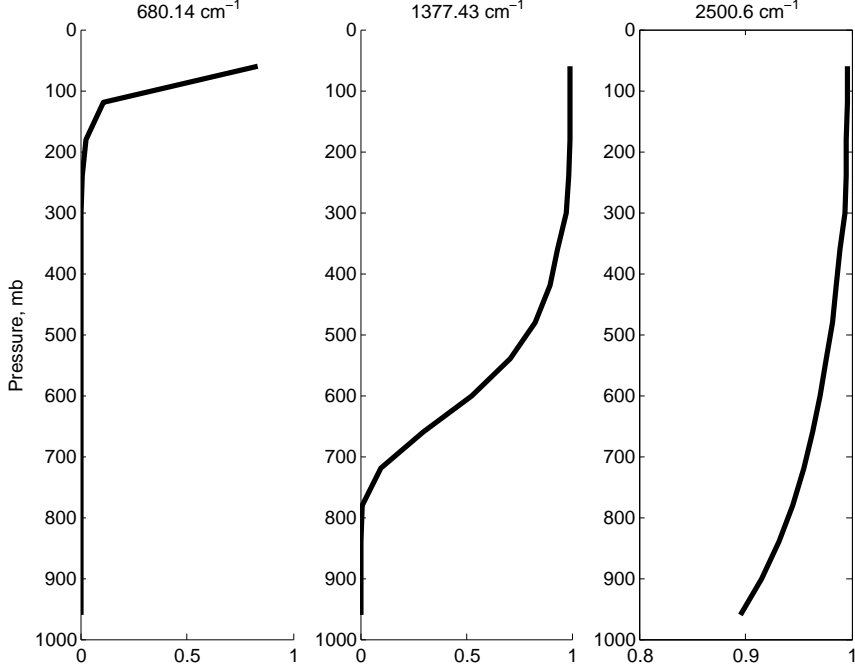


Figure 3.2: Weighting functions for three AIRS channels.

In the first line of (3.18), the terms inside the bracket but outside the integral give radiation emitted from the surface, while the integral gives radiance from below the cloud. In the second line, the first term gives the radiation emitted from the cloud and the integral gives the radiation from above the cloud.

The monochromatic transmittance in this case is defined by

$$T_\lambda(p_1, p_2) = \exp \left( -\frac{1}{g} \int_{p_1}^{p_2} \kappa_\lambda(p) q(p) dp \right).$$

Note that we have the following

$$T_\lambda(p_s, p_c) \times T_\lambda(p_c, 0) = T_\lambda(p_s, 0), \quad T_\lambda(p, p_c) \times T_\lambda(p_c, 0) = T_\lambda(p, 0). \quad (3.19)$$

Then (3.18) simplifies as

$$\begin{aligned} I_\lambda^{cld} &= (1 - \eta\epsilon_\lambda) \left[ B_\lambda(T_s) T_\lambda(p_s, 0) + \int_{p_s}^{p_c} B_\lambda[T(p)] \frac{\partial T_\lambda(p, 0)}{\partial p} dp \right] \\ &+ \eta\epsilon_\lambda B_\lambda(T_c) T_\lambda(p_c, 0) + \int_{p_c}^0 B_\lambda[T(p)] \frac{\partial T_\lambda(p, 0)}{\partial p} dp. \end{aligned} \quad (3.20)$$

Notice in (3.20) that if the weighing function

$$\frac{\partial T_\lambda(p, 0)}{\partial p}$$

peaks between the cloud top and the surface of the Earth, then the observed radiance intensity at the top of the atmosphere will be diminished by the cloud. This is because at pressures lower than  $p_c$  (at greater height), the weighting function will be small. The integral outside the bracket in (3.20) will have little contribution to the observed radiance. Depending on how large  $\eta$  is, the terms inside the bracket may also have little contribution.

We will show that clear sky brightness temperatures are larger than their corresponding cloudy sky brightness temperatures. For a clear field of view,  $\eta\epsilon_\lambda = 0$  and the satellite measured radiance is

$$I_\lambda^{clr} = B_\lambda(T_s)T_\lambda(p_s, 0) + \int_{p_s}^0 B_v[T(p)] \frac{\partial T_\lambda(p, 0)}{\partial p} dp. \quad (3.21)$$

Subtracting (3.21) from (3.20) gives

$$\begin{aligned} I_\lambda^{cld} - I_\lambda^{clr} &= -\eta\epsilon_\lambda B_\lambda(T_s)T_\lambda(p_s, 0) \\ &\quad - \eta\epsilon_\lambda \int_{p_s}^{p_c} B_\lambda[T(p)] \frac{\partial T_\lambda(p, 0)}{\partial p} dp + \eta\epsilon_\lambda B_\lambda(T_c)T_\lambda(p_c, 0). \end{aligned}$$

This is the cloud signal in the satellite measured radiances for wavelength  $\lambda$ . To simplify this equation, we integrate by parts. Let

$$u = B_\lambda[T(p)], \quad dv = \frac{\partial T_\lambda(p, 0)}{\partial p} dp$$

so that

$$du = \left( \frac{\partial B_\lambda}{\partial T} \right) \left( \frac{\partial T}{\partial p} \right) dp, \quad v = T_\lambda(p, 0).$$

Then

$$\begin{aligned} I_\lambda^{cld} - I_\lambda^{clr} &= -\eta\epsilon_\lambda B_\lambda(T_s)T_\lambda(p_s, 0) \\ &\quad - \eta\epsilon_\lambda \left( [B_\lambda[T(p)]T_\lambda(p, 0)]_{p_s}^{p_c} - \int_{p_s}^{p_c} T_\lambda(p, 0) \left[ \frac{\partial B_\lambda}{\partial T} \right] \left[ \frac{\partial T}{\partial p} \right] dp \right) \\ &\quad + \eta\epsilon_\lambda B_\lambda(T_c)T_\lambda(p_c, 0). \end{aligned} \quad (3.22)$$

Now

$$[B_\lambda[T(p)]T_\lambda(p, 0)]_{p=p_s}^{p=p_c} = B_\lambda(T_c)T_\lambda(p_c, 0) - B_\lambda(T_s)T_\lambda(p_s, 0).$$

Substituting this back into (3.22) cancels out all non-integral terms and so

$$I_{\lambda}^{cd} - I_{\lambda}^{clr} = \eta \epsilon_{\lambda} \int_{p_s}^{p_c} T_{\lambda}(p, 0) \left[ \frac{\partial B_{\lambda}}{\partial T} \right] \left[ \frac{\partial T}{\partial p} \right] dp. \quad (3.23)$$

Since  $T_{\lambda}(p, 0)$  is an exponential function, it is always positive. Differentiating

$$B_{\lambda}(T) = \frac{2hc^2 \lambda^{-5}}{e^{hc/k\lambda T} - 1},$$

we also have

$$\frac{\partial B_{\lambda}}{\partial T} = \frac{2h^2 c^3 \lambda^{-5} e^{hc/k\lambda T}}{k\lambda T^2 (e^{hc/k\lambda T} - 1)^2} > 0.$$

Finally, if we assume the ideal gas law,  $p = \rho_a R_a T$ , where  $R_a$  is the gas constant associated with the absorbing gas, we have

$$\frac{\partial T}{\partial p} = \frac{1}{\rho_a R_a} > 0.$$

Thus the integrand in (3.23) is positive. However, in the limits of integration,  $p_c < p_s$ , meaning that this integral is negative. This means that the observed intensity over a cloudy field of view will always be less than or equal to the observed intensity if the field of view were clear.

### 3.3 Interpolating Multicloud Model Variables to RTM Radiances

The Prototype Community Radiative Transfer Model (pCRTM), [22], is a model that simulates satellite observations by using the radiative transfer model. This model uses known properties of the atmospheric state, such as temperature and humidity, to compute the brightness temperature at the top of the atmosphere. The pCRTM can be configured to simulate observations from different satellite instruments, such as AIRS or Moderate-Resolution Imaging Spectroradiometer (MODIS). The pCRTM has  $L$  vertical discrete layers  $z_1, \dots, z_L$  where  $z_1$  corresponds to the layer at the top of the atmosphere, bounded by  $p = 0$  mb, and where  $z_L$  denotes the surface layer, whose edge is located at pressure  $p_s = 1,000$  mb. The vertical height of  $z_i$  falls in the middle of the  $i^{\text{th}}$  atmospheric layer. In our experiments, we use a vertical discretization of  $L = 16$  atmospheric layers to correspond to the tropospheric height  $H_T = 16$  km of the multicloud model.

To generate satellite observations, we use a numerical simulation of the multicloud model as the underlying true atmospheric state. Since the pCRTM takes temperature,  $T(p)$  (in K), water vapor,  $q_v(p)$  (in g/kg), and surface temperature,  $T_s$  (in K), as inputs, we need to extrapolate the solution of the multicloud model to these units at appropriate pressure coordinate discrete

points. Of the eight multicloud prognostic variables, we use  $\theta_1$  and  $\theta_2$ , the potential temperature of each mode (in K),  $\theta_{eb}$ , the equivalent potential temperature anomaly at the boundary layer (in K, at height 500 meters), and  $q$ , the vertically averaged moisture level (in K) to derive the temperature and moisture in each atmospheric layer and the surface temperature. In the remainder of this section, we derive the physical relations between the output of the multicloud model and the inputs to the pCRTM.

We first need to convert between height and pressure coordinates. Hydrostatic balance is given by

$$\frac{\partial p}{\partial z} = -g\rho$$

where  $\rho$  is the density of the atmosphere. Using the ideal gas law  $p = \rho RT$  gives

$$\frac{\partial p}{\partial z} = -\frac{gp}{RT}.$$

The temperature  $T$  has a dependence on  $z$  and we assume the piecewise linear temperature profile from [58]

$$T = \begin{cases} mz + T_o & \text{if } z < 11 \text{ km} \\ T_o + 11m & \text{if } z \geq 11 \text{ km,} \end{cases}$$

with constants  $m$  and  $T_o$  (Note that this equation only holds for  $z \leq 16$  km). We integrate for each case. First, let  $z < 11$  km. Then

$$\frac{\partial p}{\partial z} = -\frac{gp}{R(mz + T_o)}$$

which we can solve via separation of variables.

$$\begin{aligned} \int \frac{dp}{p} = \ln p &= -\frac{g}{R} \int \frac{dz}{mz + T_o} \\ &= -\frac{g}{mR} \ln(mz + T_o) + C \\ &= \ln(mz + T_o)^{-g/mR} + C. \end{aligned}$$

Then

$$z = kp^{-\frac{mR}{g}} - \frac{T_o}{m}.$$

If  $p = p_o$  when  $z = 0$  then

$$k = \frac{T_o}{m} p_o^{\frac{mR}{g}}$$

and

$$z = \frac{T_o}{m} \left[ \left( \frac{p_o}{p} \right)^{\frac{Rm}{g}} - 1 \right].$$

Now suppose that  $z \geq 11$  km. At this height, the temperature profile is assumed to be constant. For simplicity, define  $T_1 = mz + T_o$ . Then the hydrostatic equation becomes

$$\frac{\partial p}{\partial z} = -\frac{gp}{RT_1}$$

which has solution

$$p = k \exp \left( -\frac{gz}{RT_1} \right).$$

If  $p = p_o$  when  $z = 0$  then  $k = p_o$  and

$$\begin{aligned} z &= -\frac{RT_1}{g} \ln \left( \frac{p}{p_o} \right) \\ &= \frac{R(11m + T_o)}{g} \ln \left( \frac{p_o}{p} \right) \end{aligned}$$

Putting both cases together, the relationship between pressure  $p$  and height  $z$  is

$$z = \begin{cases} \frac{T_o}{m} \left[ \left( \frac{p_o}{p} \right)^{\frac{Rm}{g}} - 1 \right] & \text{if } z < 11 \text{ km} \\ \frac{R(11m + T_o)}{g} \ln \left( \frac{p_o}{p} \right) & \text{if } z \geq 11 \text{ km,} \end{cases}$$

where  $p_o$  is the pressure at height  $z = 0$ .

The multicloud variables  $\theta_1$  and  $\theta_2$  have no height dependence. To build a vertical profile, we use

$$\Theta'(x, z, t) = G'(z)\theta_1(x, t) + 2G'(2z)\theta_2(x, t),$$

which follows from (3.2), where  $G'(z) = \sqrt{2} \sin(\pi z/H_T)$  and  $H_T$  is the height of the troposphere [32]. The term  $\Theta'$  represents a perturbation from a mean potential temperature  $\bar{\Theta}$  so that the vertical potential temperature is given by

$$\Theta(x, z, t) = \bar{\Theta} + \Theta'(x, z, t).$$

The relationship between temperature and potential temperature is given by

$$T = \Theta \left( \frac{p}{p_o} \right)^{\frac{R}{C_p}} \quad (3.24)$$

where  $R$  is the gas constant of dry air and  $C_p$  is the heat capacity of dry air at constant

temperature [65].

In terms of the multcloud model,  $q$  gives the vertically averaged moisture content. The pCRTM model incorporates height-dependent moisture. Let us write the total vertical moisture content  $q_v$  as

$$q_v(x, z, t) = Q(z) + \tilde{q}(x, z, t),$$

and assume that we can decompose  $\tilde{q}$  as

$$\tilde{q}(x, z, t) = q_1(z)q_2(x, t).$$

If we compute the vertical average (denoted in this section by  $\langle f \rangle = (1/H_T) \int_0^{H_T} f(z)dz$ ), we obtain

$$\langle \tilde{q} \rangle = \langle q_1 \rangle q_2,$$

since  $q_2$  has no dependence on  $z$ . Solving this equation for  $q_2$ , we obtain

$$\tilde{q}(x, z, t) = \frac{q_1(z)}{\langle q_1 \rangle} \langle \tilde{q} \rangle.$$

The average  $\langle \tilde{q} \rangle$  is equal to the multcloud variable  $q$ . Furthermore, we will assume that  $Q(z)$  and  $q_1$  are exponential, such that

$$q_1 = \exp\left(\frac{-z}{H_q}\right), \quad Q(z) = q_o q_1,$$

where  $H_q$  gives the vertical scale above which moisture is negligible and  $q_o$  is the surface moisture. With these exponential expressions, the total moisture is given by

$$q_v(x, z, t) = \left(q_o + \frac{\langle \tilde{q} \rangle}{\langle q_1 \rangle}\right) \exp\left(\frac{-z}{H_q}\right).$$

We can write this in pressure coordinates by substituting  $z(p)$ :

$$q_v(x, p, t) = \left(q_o + \frac{\langle \tilde{q} \rangle}{\langle q_1 \rangle}\right) \exp\left(\frac{T_o}{mH_q} \left(1 - \left(\frac{p_o}{p}\right)^{\frac{Rm}{g}}\right)\right).$$

Lastly, we convert to units of g/kg, expressing a mixing ratio, ( $\langle \tilde{q} \rangle = q$  has units of Kelvin), by multiplying  $\langle \tilde{q} \rangle$  by  $C_p/L_v$ . The relation between the  $q$  of the multcloud model and the  $q_v$  required by pCRTM is

$$q_v = \left(q_o + \frac{C_p q}{L_v H_q (1 - e^{-1})}\right) \exp\left(\frac{T_o}{mH_q} \left(1 - \left(\frac{p_o}{p}\right)^{\frac{Rm}{g}}\right)\right), \quad (3.25)$$

where  $C_p$  is the heat capacity of water vapor at constant pressure and  $L_v$  is the latent heat of vaporization, of water. This is the equation for  $z < 11$  km. For  $z \geq 11$  km we have

$$q_v = \left(\frac{p}{p_o}\right)^{\frac{R(11m+T_o)}{gH_q}} \left[ q_o + \frac{C_p q}{L_v H_q (1 - e^{-1})} \right]. \quad (3.26)$$

Assuming constant  $T_s$  and Latent heat of vaporization  $L_v$ , the equivalent potential temperature at the boundary is

$$\begin{aligned} \theta_{eb} &= \theta \exp\left(\frac{L_v q_o}{C_p T_s}\right) \\ &= T_s \left(\frac{p_o}{p}\right)^{\frac{R}{C_p}} \exp\left(\frac{L_v q_o}{C_p T_s}\right), \end{aligned}$$

with (3.24) [65]. Expanding the left hand side and the exponential term on the right side as a Taylor series gives the simplification

$$\bar{\theta}_{eb} + \theta'_{eb} \approx \left(\frac{p_o}{p}\right)^{\frac{R}{C_p}} \left(T_s + \frac{L_v q_o}{C_p}\right)$$

where  $\theta'_{eb} = \theta_{eb}$  is outputted from the multicloud model.

Therefore, we interpolate the multicloud model variables  $\theta_1$ ,  $\theta_2$ ,  $\theta_{eb}$  and  $q$  into temperature  $T$ , sea surface temperature  $T_s$  and water vapor  $q_v$  through the following equations, where pressure  $p$  is given by

$$p = \begin{cases} p_o \left(\frac{mz}{T_o} + 1\right)^{-\frac{g}{Rm}} & \text{if } z < 11 \text{ km} \\ p_o \exp\left(-\frac{gz}{R(11m+T_o)}\right) & \text{if } z \geq 11 \text{ km.} \end{cases}$$

The temperature  $T$  in terms of potential temperature  $\theta$  is given by

$$T(p) = (\bar{\Theta} + G'(z(p))\theta_1 + 2G'(2z(p))\theta_2) \left(\frac{p}{p_o}\right)^{\frac{R}{C_p}}.$$

The water vapor  $q_v$  in terms of the vertically averaged water vapor  $q$  is given by (3.25)-(3.26). The sea surface temperature  $T_s$  is given by

$$T_s = (\bar{\theta}_{eb} + \theta'_{eb}) \left(\frac{p}{p_o}\right)^{\frac{R}{C_p}} - \frac{L_v q_o}{C_p}.$$

The constants used in these relations are given in Table 3.2.



Table 3.2: The values of constants used to interpolate multicloud model variables.

Parameter	Value
$C_p$ air	1005.7 J / kg K ( $\pm 2.5$ )
$C_p$ water	1870 J / kg K ( $\pm 25$ )
$g$	9.8 m/s
$H_q$	2200 m
$H_T$	16000 m
$L_v$	$L_o - L_1(T - 273.15)$ J / g
$L_o$	2501 J / g
$L_1$	2.32 J / g K ( $\pm 0.55$ )
$m$	-0.0065 K / m, [58]
$p_o$	1000 mb
$q_o$	20 g / kg
$R$	287.04 J / kg K
$T_o$	303 K
$\Theta$	322 K
$\theta_{eb}$	330 K

### 3.3.1 Adding Cloud Cover to Synthetic Satellite Observations

We now discuss how to simulate cloudy AIRS data. Recall that the cloud free (or clear) observed radiance of wavelength  $\lambda$  is

$$I_\lambda^{clr} = B_\lambda(T_s)T_\lambda(p_s) + \int_{p_s}^0 B_\lambda[T(p)] \frac{\partial T_\lambda}{\partial p} dp \quad (3.27)$$

where  $p_s = 1,000$  mb. For simplicity, let  $c$  denote the percent of cloud cover ( $c = \eta\epsilon$ ) and let the cloud top be located at pressure  $p_c$ . We can assume the cloud is a thin sheet. Then in the presence of clouds, the observed radiance is

$$\begin{aligned} I_\lambda^{cld} &= (1 - c) \left[ B_\lambda(T_s)T_\lambda(p_s, 0) + \int_{p_s}^{p_c} B_\lambda(p) \frac{\partial T_\lambda(p, 0)}{\partial p} dp \right] \\ &+ c B_\lambda(T_c)T_\lambda(p_c, 0) + \int_{p_c}^0 B_\lambda(p) \frac{\partial T_\lambda(p, 0)}{\partial p} dp. \end{aligned} \quad (3.28)$$

The main issue lies in determining  $c$  and  $p_c$ . For simplicity we assume that there is only one cloud that affects the computation of  $I_\lambda^{cld}$ : a low lying congestus cloud, a higher stratiform cloud or a deep convective cloud. For a congestus cloud, we will take  $p_c$  to be the pressure at height 2 km, and for both the stratiform and deep convective clouds, we will take  $p_c$  to be

the pressure at 12 km. The value of  $c$  will depend on  $H_c$ ,  $H_s$  and  $P$ , as well as the switching function  $\Lambda$  defined in (3.8), since these variables correlate with the type of cloud present. In the top panel of Figure 3.3, we plot the true value of  $\Lambda$  at one location as a function of time, while in the bottom panel, we plot the normalized true values of  $P$ ,  $H_s$  and  $H_c$  at the same location, over time. We normalize  $P$  and  $H_s$  by dividing each by its respective overall maximal value. To normalize  $H_c$ , we first subtract its overall minimum value, before dividing by its new maximal value. Notice here that the normalized stratiform heating rate and precipitation rate exhibit nearly the same behavior. Since both cloud top heights are assumed to be the same, we will treat the deep convection heating (precipitation) and the stratiform heating as one type of cloud in our simulations.

Notice in Figure 3.3 that a local maximum of  $\Lambda$  corresponds to a local maximum in the normalized value of  $H_c$ . Therefore, in our simulations, we set the cloud type to be congestus when  $\Lambda > 0.7$ . Moreover, we take the cloud fraction  $c$  to be normalized value of  $H_c$ . The precipitation rate is near a local maximum when  $\Lambda$  is near a local minimum. Here, we set the cloud type to be the deep convective cloud when  $\Lambda < 0.5$ , and take  $c$  to be the normalized value of  $P$ . When  $0.5 \leq \Lambda \leq 0.7$ , we assume that there is no cloud present, and set  $c = 0$  and  $p_c = 0$ . To summarize, we determine the cloud cover percent and pressure with the following:

$$(c, p_c) = \begin{cases} \left( \frac{P}{\max\{P\}}, p(12 \text{ km}) \right), & \text{if } \Lambda < 0.5 \\ \left( \frac{H_c - \min\{H_c\}}{\max\{H_c - \min\{H_c\}\}}, p(2 \text{ km}) \right), & \text{if } \Lambda > 0.7 \\ (0, 0), & \text{if } 0.5 \leq \Lambda \leq 0.7. \end{cases} \quad (3.29)$$

Figure 3.4 shows brightness temperatures calculated over a clear field of view in the left panels for three different channels at a particular location (the weighting functions of these channels are shown in Figure 3.2). The right panels show brightness temperatures for these same three channels and location, but after cloud cover was added to the RTM. For the higher frequency channels, cloudy sky brightness temperatures are consistently lower than clear sky radiances. In the lowest frequency channel, the brightness temperatures are roughly the same for both cases. This is because the weighting function for this channel peaks at the top of the atmosphere, (see the first panel of Figure 3.2) and the observed radiances emanate from the atmospheric layer above cloud tops. Notice that the cloudy sky brightness temperatures associated with wavenumber  $1377.426 \text{ cm}^{-1}$  agrees with the analogous clear sky brightness temperatures at several times. This channel is sensitive to cloud effects, however, it is not sensitive to the effects of low lying congestus clouds; its weighting function is nearly zero below 2 km (see the middle panel of Figure 3.2). For this channel, deep spikes in the cloudy sky temperatures can be explained by the presence of a deep convective cloud. The channel associated with wavenumber  $2500.602 \text{ cm}^{-1}$  is sensitive to all cloud types, because its weighting function weights over the

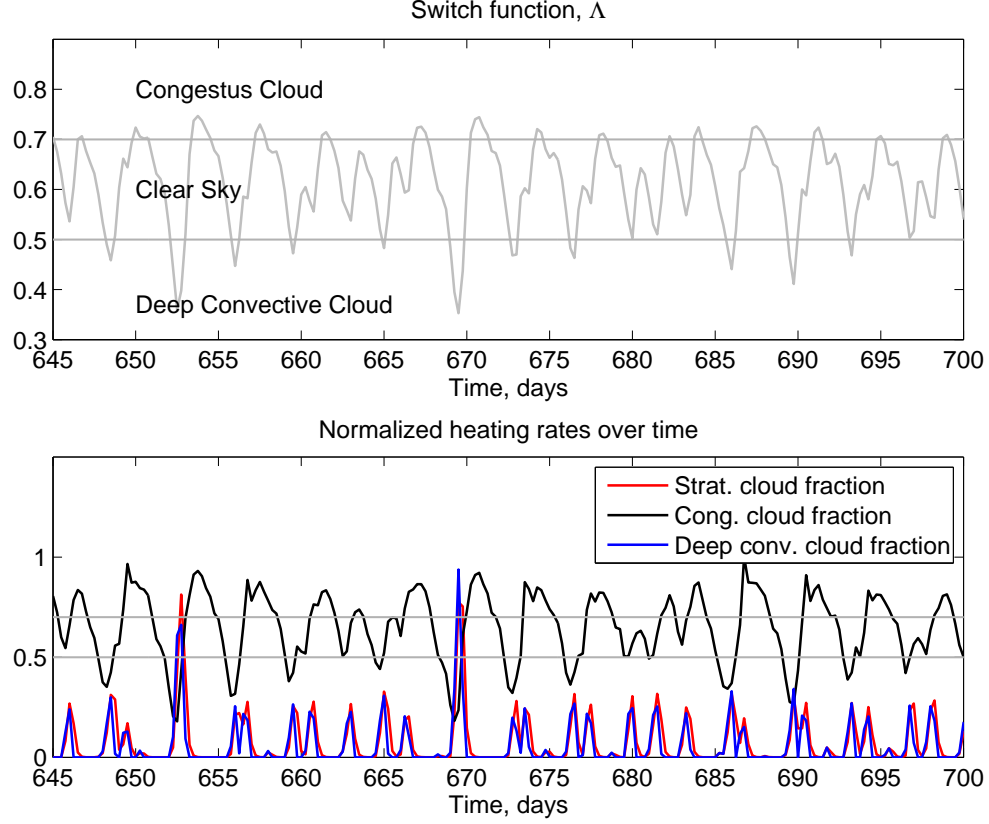


Figure 3.3: The switch function  $\Lambda$  (top panel) and the normalized deep convective, stratiform, and congestus heating rates, which correspond to the cloud fraction (bottom panel), at location  $x = 20,000$  over time.

entire atmospheric column (see the right panel of Figure 3.2). Therefore, for this channel, there is a large difference between clear and cloudy brightness temperature.

Next, we examine the brightness temperature at one particular location at times that each of these clouds appear. Figure 3.5 shows the brightness temperature measured by each of the 281 channels at a time when each cloud type is present. Notice that the clear sky temperatures in Figure 3.5 are higher than the cloudy sky temperatures. In the case of a congestus cloud, cloudy cover is  $c = 56\%$ . This cloud type does not have a great vertical extent, so several channels observe nearly the same temperature in both the clear and cloudy cases. In the case of a deep convective cloud, with cloud cover  $c = 100\%$ , cloudy sky radiances are severely diminished by the cloud.

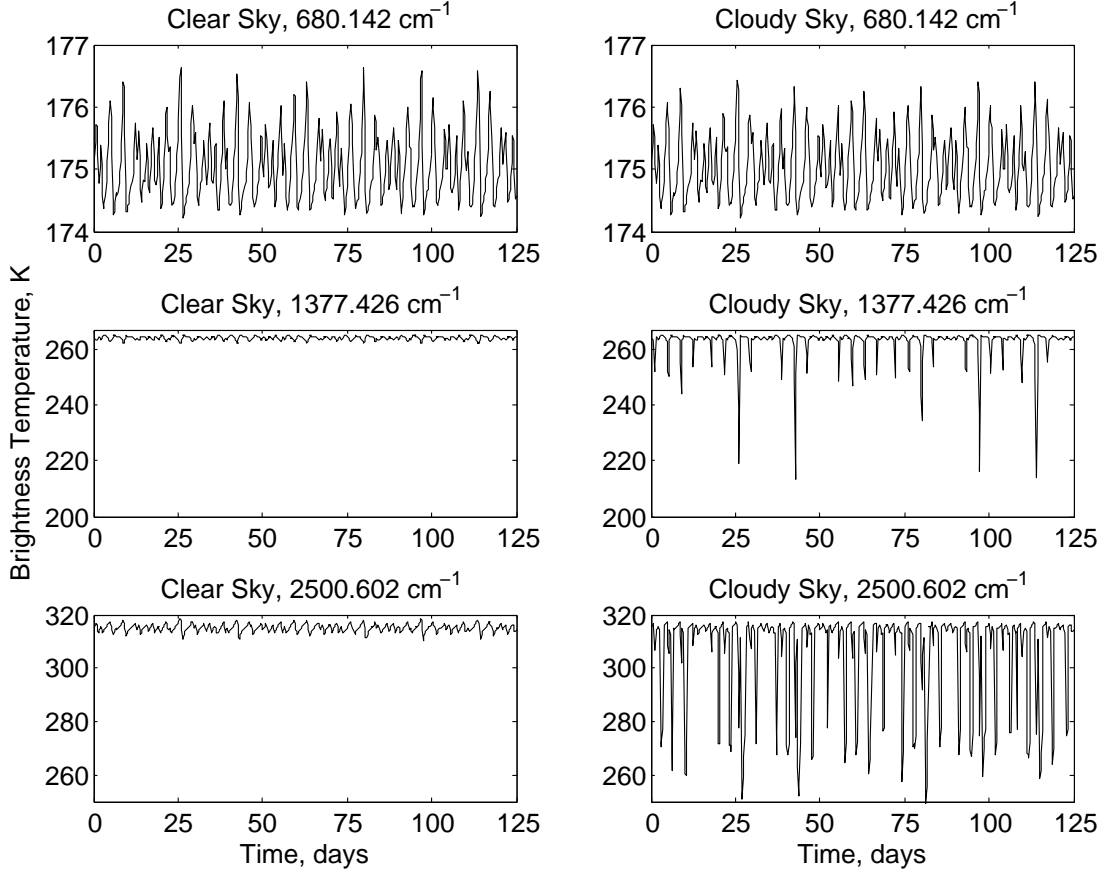


Figure 3.4: Clear (left) and cloudy (right) AIRS observations at one location, over time. The top panels correspond to channel frequency  $680.142 \text{ cm}^{-1}$ , the middle panels correspond to channel frequency  $1377.426 \text{ cm}^{-1}$  and the bottom panels correspond to channel frequency  $2500.602 \text{ cm}^{-1}$ .

### 3.4 Data Assimilation

The data assimilation flow is illustrated in Figure 3.6. It starts with the selection of the initial ensemble members, chosen at random times of the multicloud model true solution. One ensemble member includes the estimates of the eight multicloud model variables  $u_1$ ,  $u_2$ ,  $\theta_1$ ,  $\theta_2$ ,  $\theta_{eb}$ ,  $q$ ,  $H_s$ , and  $H_c$  over  $M = 1,000$  locations, equally spaced along the equator. We denote the prior state estimate

$$x^b = (u_1, u_2, \theta_1, \theta_2, \theta_{eb}, q, H_s, H_c)^b,$$

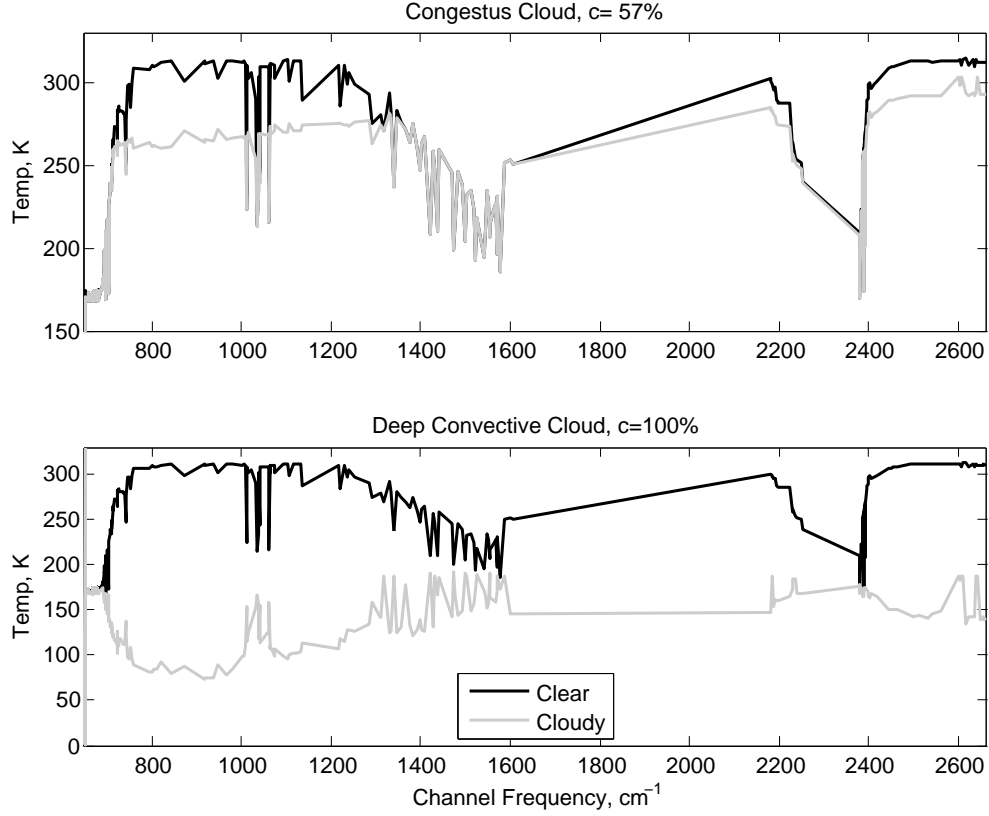


Figure 3.5: Congestus cloud (top), stratiform cloud (middle) and deep convective cloud (bottom) effects on brightness temperatures for all channels. The black lines represent the clear sky pCRTM radiances while gray lines represent the cloudy sky radiances.

the transformed prior state estimate  $\xi^b$  and the forward RTM  $h(\xi^b)$ , which computes the brightness temperature at each spatial location. The transformed estimates  $\xi^b$  represent the result of extrapolating the temperature and humidity estimates to different layers of the atmosphere (see Section 3.3). The synthetic noisy observations of brightness temperature and  $h(\xi^b)$  are input into the data assimilation algorithm. The posterior state estimate  $x^a$  is output from the data assimilation algorithm. Then  $x^a$  is used as an initial condition in the multicloud model to compute the new prior state estimate  $x^b$  at the next time step.

### 3.4.1 Local Ensemble Transform Kalman Filter

We will use the local ensemble transform Kalman filter (LETKF), developed by [6], in the data assimilation step. Ensemble Kalman filters use an ensemble of state vectors to represent the

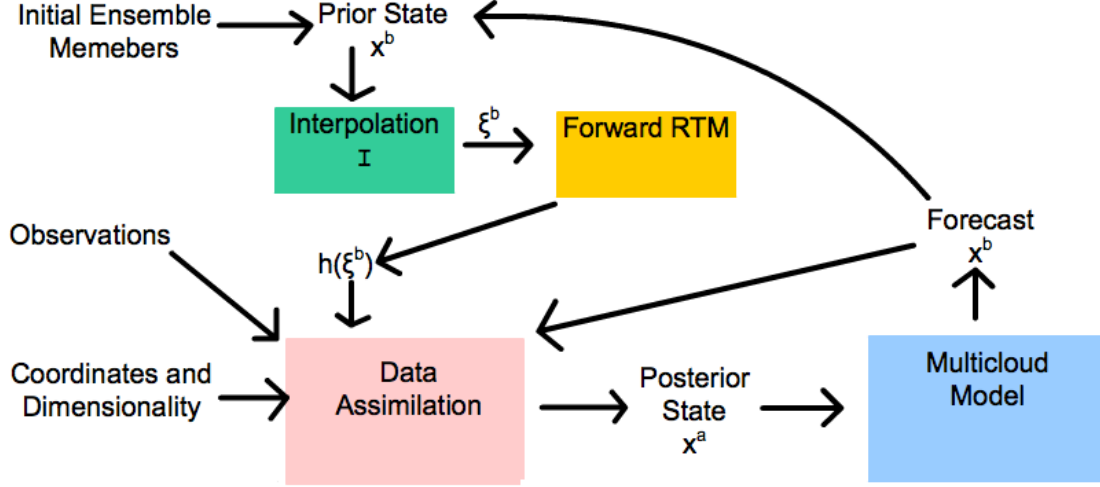


Figure 3.6: Data assimilation of satellite observations with the multicloud model.

distribution of the true state. Each ensemble member is propagated forward in time individually by the forecast model (here the forecast model is the multicloud model). Let  $\{x_k^b\}_{k=1}^K$  denote this forecast ensemble, which has size  $K$ . When observations become available, the filter uses the forecast (prior) ensemble mean,

$$\bar{x}^b = \frac{1}{K} \sum_{k=1}^K x_k^b$$

as an estimate of the atmospheric state, before accounting for the observations.

We define the brightness temperature observation at wavelength  $\lambda$ ,  $v_\lambda$ , as a function of the multicloud prognostic variables  $\xi(x) = \xi(\theta_1, \theta_2, \theta_{eb}, q)$ , by

$$v_\lambda = h(\xi(x)), \quad (3.30)$$

where the operator  $\xi$  incorporates the numerous dimensionalization transformations defined in Section 3.3 and the observation operator  $h$  is composite of the Planck function  $B_\lambda(T)$  and the radiative transfer model.

The ensemble Kalman filter uses the prior ensemble  $\{x_k^b\}_{k=1}^K$ , the observations  $v_\lambda$  and their error covariance, and the nonlinear composite  $h \circ \xi$  to calculate the analysis (posterior) ensemble mean  $\bar{x}^a$  and error covariance  $R^a$ . The forecast error covariance  $R^b$  depends on the perturbation

of each forecast ensemble member about the prior mean, and is defined by

$$R^b = \frac{1}{K-1} X^b (X^b)^T, \quad (3.31)$$

where the  $k^{th}$  column of  $X^b$  is given by  $x_k^b - \bar{x}^b$ . The key idea of the LETKF is to use a transformation  $X^a = X^b T$  so that the perturbations describe the posterior error covariance after updating locally with available observations. The perturbations in  $X^b$  are weighted such that the  $k^{th}$  column of  $X^a$  is  $x_k^a - \bar{x}^a$ . Therefore, the posterior error covariance is given by

$$R^a = \frac{1}{K-1} (X^b T) (X^b T)^T = \frac{1}{K-1} X^a (X^a)^T.$$

Each ensemble member updates based on its corresponding transformed perturbation.

The goal of the standard Kalman filter is to minimize the cost function

$$J(x) = (x - \bar{x}^b)^T (R^b)^{-1} (x - \bar{x}^b) + (v_\lambda - h(\xi(x)))^T (R^o)^{-1} (v_\lambda - h(\xi(x))),$$

where  $R^o$  gives the observation error covariance. If the observation operator  $h \circ \xi$  is linearized as  $H$ , then  $J(x)$  is minimized by

$$\bar{x}^a = \bar{x}^b + \mathcal{K} (v_\lambda - H \bar{x}^b),$$

where the Kalman gain matrix  $\mathcal{K}$  is given by

$$\mathcal{K} = ((R^b)^{-1} + H^T (R^o)^{-1} H)^{-1} H^T (R^o)^{-1}.$$

In terms of the Kalman gain, the posterior error covariance is given by

$$R^a = (\mathcal{I} - \mathcal{K} H) R^b. \quad (3.32)$$

Notice in the right side of  $\mathcal{K}$  that

$$H^T (R^o)^{-1} (H R^b H^T + R^o) = ((R^b)^{-1} + H^T (R^o)^{-1} H) R^b H^T.$$

This implies

$$\mathcal{K} = R^b H^T (H R^b H^T + R^o)^{-1}.$$

Using the expression for  $R^b$  in (3.31), the Kalman gain matrix becomes

$$\begin{aligned}
\mathcal{K} &= R^b H^T (H R^b H^T + R^o)^{-1} \\
&= \frac{1}{(K-1)} (X^b (X^b)^T) H^T \left( \frac{1}{(K-1)} H (X^b (X^b)^T) H^T + R^o \right)^{-1} \\
&= X^b (H X^b)^T (H X^b (H X^b)^T + (K-1) R^o)^{-1}
\end{aligned}$$

Observe in this last expression that when the linearized operator  $H$  appears, it acts on the perturbation matrix  $X^b$ . Therefore, rather than linearizing the complicated  $h \circ \xi$ , we can use a perturbation matrix  $V$  given by

$$V = [\vec{h}(\xi_1^b) - \vec{v}; \vec{h}(\xi_2^b) - \vec{v}; \dots; \vec{h}(\xi_K^b) - \vec{v}],$$

where  $\vec{v} = \overline{\vec{h}(\xi^b)}$ . Then,

$$\begin{aligned}
\mathcal{K} &= X^b V^T (V V^T + (K-1) R^o)^{-1} \\
&= X^b (V^T (R^o)^{-1} V + (K-1) \mathcal{I})^{-1} V^T (R^o)^{-1} \\
&= X^b J^{-1} V^T (R^o)^{-1},
\end{aligned} \tag{3.33}$$

where

$$J = (K-1)I + V^T (R^o)^{-1} V. \tag{3.34}$$

When an observation  $v_\lambda$  becomes available, the posterior ensemble mean is updated with

$$\bar{x}^a = \bar{x}^b + \mathcal{K}(v_\lambda - \vec{h}(\xi^b)). \tag{3.35}$$

Notice in (3.35) that if the observations have large noise then the norm of  $\mathcal{K}$  is small and therefore the filter will tend to trust the model dynamics. On the other hand, if the observations have small noise, the norm of  $\mathcal{K}$  will be larger and the filter will tend to trust the observations.

To compute the transformation matrix  $T$ , we use (3.31) and (3.33) in the posterior error



covariance expression in (3.32):

$$\begin{aligned}
R^a &= (\mathcal{I} - \mathcal{K}H)R^b \\
&= \left( \mathcal{I} - X^b(V^T(R^o)^{-1}V + (K-1)\mathcal{I})^{-1}V^T(R^o)^{-1}H \right) \left( \frac{X^b(X^b)^T}{K-1} \right) \\
&= \frac{1}{K-1}X^b \left( \mathcal{I} - \left( \frac{V^T(R^o)^{-1}V}{K-1} + \mathcal{I} \right)^{-1} \left( \frac{V^T(R^o)^{-1}V}{K-1} \right) \right) (X^b)^T \\
&= \frac{1}{K-1}X^b (\mathcal{I} - (A + \mathcal{I})^{-1}A) (X^b)^T \\
&= \frac{1}{K-1}X^b(\mathcal{I} + A)^{-1}(X^b)^T
\end{aligned}$$

where

$$A = \frac{V^T(R^o)^{-1}V}{K-1}.$$

Thus

$$\begin{aligned}
R^a &= X^b((K-1)\mathcal{I} + V^T(R^o)^{-1}V)^{-1}(X^b)^T \\
&= X^b J^{-1}(X^b)^T \\
&= \frac{X^b T T^T (X^b)^T}{K-1} \\
&= \frac{X^b T (X^b T)^T}{K-1},
\end{aligned}$$

where  $J$  is given in (3.34).

Notice that the columns of  $V$  sum to the zero vector, implying that  $J$  is not full rank. Therefore, to compute the transformation matrix  $T$ , we compute the singular value decomposition of  $J$  as

$$J = Y\Gamma Y^T.$$

Then the transformation matrix  $T$  is given by

$$T = \sqrt{K-1}Y\Gamma^{-1/2}Y^T.$$

This choice of  $T$  permits a symmetric transformation. (In [66], this type of LETKF is called the spherical simplex LETKF). Each posterior ensemble member  $x_k^a$ ,  $1 \leq k \leq K$  is then obtained by adding the posterior mean  $\bar{x}^a$  to the  $k^{th}$  column of  $X^a = X^b T$ .

Because we are under-sampling the atmospheric state distribution by using  $K$  ensemble members, we inflate  $X^b$  and  $V$  by an inflation percent  $r$ , by multiplying these matrices by a factor of  $\sqrt{1+r}$ . The value of  $r$  is determined empirically, and partially depends on the sparsity

of the observation network, the observational time length, and the ensemble size. This strategy is referred to as multiplicative variance inflation in [68, 3].

In our simulations, the forecast model is discretized by 1,000 grid points, equally spaced 40 km apart along the equator, which has length 40,000 km. The multicloud model produces a forecast every 6 hours. We implement the data assimilation algorithm locally in space, meaning that, to correct at each model grid point, we use the available observations within a box of radius  $\mathcal{D}$  grid points. Therefore,  $X^a, X^b \in \mathbb{R}^{8(2\mathcal{D}+1) \times K}$ , since there are 8 multicloud prognostic variables.

### 3.4.2 Biases from Cloudy Sky Observations

One can easily induce errors or biases in the observation operator in (3.30) if the observation model and/or its parameters are not completely specified. Previously, [16] estimated the effects of such biases in the case of clear sky AIRS measurements. A cloudy field of view can have a detrimental effect on satellite observations and we will study the potential biases that can arise in this situation. In particular, we will examine the sensitivity of the filter analysis to the error in the cloud cover fraction  $c$  defined in (3.29), and the error in  $\Lambda$ , which specifies cloud type and height. We will see that when  $c$  cannot be accurately known, we must carefully select AIRS channels for assimilation, based on the cloud height. In reality, Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) can provide accurate estimates of cloud top height [64]. In the remainder of this section, we describe how we choose cloud height-dependent channels. The channels we select in this case will have weighting functions which are nearly zero below the cloud top, so that the Planck function is only weighted above the cloud. Therefore, the brightness temperatures associated with these channels will be minimally unobstructed.

The congestus cloud has height 2,000 m, which is relatively close to the surface. In Figure 3.7, we show two AIRS channels that detect radiation of frequencies  $1036.5 \text{ cm}^{-1}$   $1377.43 \text{ cm}^{-1}$ . The shaded gray area represents the part of the atmosphere below the congestus cloud top. Notice here that the weighting function for both channels is close to zero below the cloud. This demonstrates that these two channels are insensitive to the atmospheric dynamics below the cloud top. Therefore, when assimilating cloud-dependent channels, we will use these two when a congestus cloud is present.

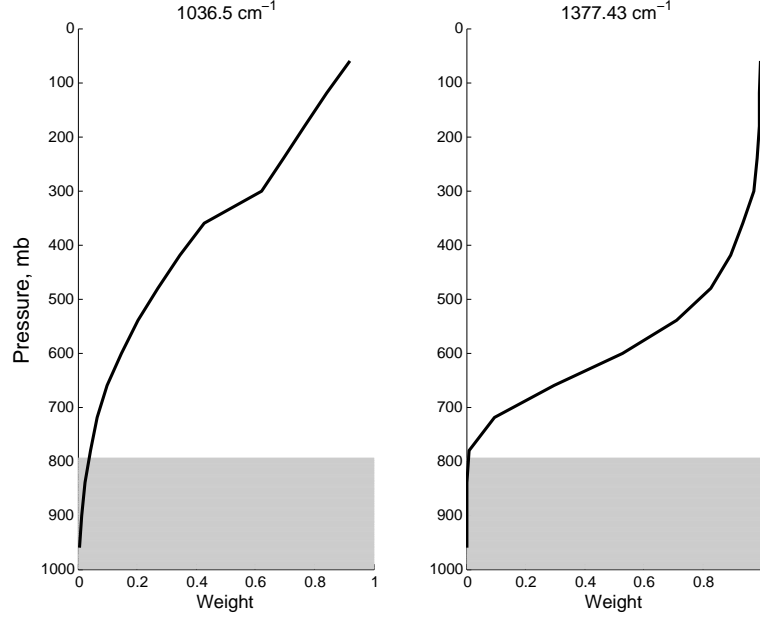


Figure 3.7: Weighting functions of AIRS channels that are minimally obstructed by congestus clouds. The gray area represents the atmospheric region below a congestus cloud.

The stratiform cloud and the deep convective cloud both have height 12,000 m, which is much closer to the top of the troposphere. Therefore, the channels that are insensitive to either of these clouds will have weighting functions that peak at the top of the atmosphere. The two channels we consider channels that detect radiation of frequencies  $680.14 \text{ cm}^{-1}$  and  $681.72 \text{ cm}^{-1}$ , depicted in Figure 3.8. The shaded gray area represents the part of the atmosphere below either cloud top. While such channels are not sensitive to the effects of clouds, they cannot observe a large part of the atmosphere, and therefore may be uninformative. Channels like these are more likely to be assimilated with current weather forecast models [54].

Finally, if the field of view, we will use the two AIRS channels associated with radiation of frequencies  $2500.6 \text{ cm}^{-1}$  and  $2561.13 \text{ cm}^{-1}$ . As illustrated in Figure 3.9, these channels observe the entire atmosphere. Because the weighting functions have large weight over the entire atmospheric column, their respective channels are extremely sensitive to cloud effects. Therefore, channels such as these are rarely used in numerical weather prediction [54].

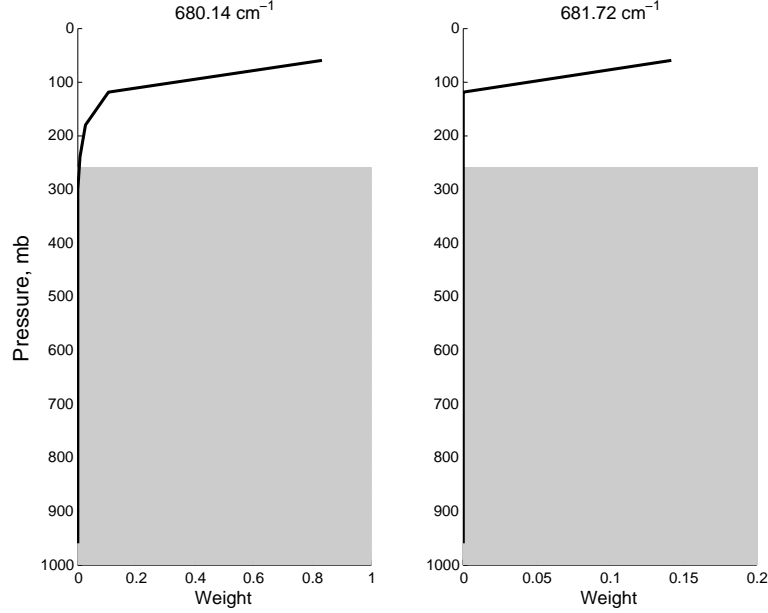


Figure 3.8: Weighting functions of AIRS channels that are minimally obstructed by deep convective clouds and stratiform clouds. The gray area shows the atmospheric region below their tops.

### 3.5 Results

We report the numerical results of several filtering experiments, after assimilating synthetic clear and cloudy sky AIRS observations, as well as direct observations of the multicloud temperature and humidity prognostic variables. We present the results of experiments that use the empirically tuned covariance inflation  $r$  in each case. For clear sky observations, the covariance inflation tends to be between 12-16 % and for cloudy sky observations, it tends to be between 4-18 %. We numerically generate each observation of brightness temperature with noise of variance 1 Kelvin, which is consistent with the temperature accuracy of AIRS [38].

In our discussions, we define  $D_{obs}$  as the distance between two observations. If we observe at every model grid point (resolved at 1,000 regularly spaced locations along the equator, with a circumference of 40,000 km) then  $D_{obs} = 40$  km. On the other hand, if we observe at every other model grid point, at 500 equally spaced locations, then  $D_{obs} = 80$  km. Finally, if we observe at every fourth model grid point, at 250 equally distributed locations, then  $D_{obs} = 160$  km. We also define the observational time length  $T_{obs}$  as the time between observational updates in the filter. Finally, we define the localization radius in terms of physical distance, in kilometers. Therefore, a radius of  $\mathcal{D} = 1$  observation corresponds to a localization radius of 40 km; a radius

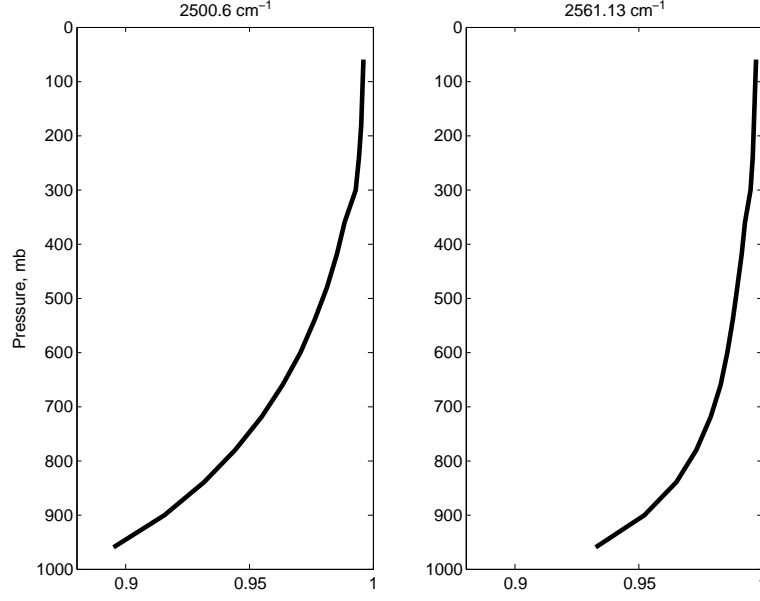


Figure 3.9: Weighting functions of AIRS channels that observe the entire atmosphere.

of  $\mathcal{D} = 2$  observations corresponds to a localization radius of 80 km.

### 3.5.1 Direct Observations

We first show the results of assimilating only direct observations of the temperature variables  $\theta_1$ ,  $\theta_2$  and  $\theta_{eb}$  and of the humidity variable  $q$  in the filtering experiment. Here, we perturb an observation of each variable with noise of variance equal to 10 percent of its overall variance. The numerical results that we obtain with these observations can be considered to be the theoretically best estimates that can be achieved by satellite measurement, since the RTM uses these variables to generate brightness temperatures. In Figures 3.10 and 3.11 we show an example of the posterior and prior state estimates of each multicloud prognostic variable, and the precipitation rate, compared with the true signal, using observations at each of the 1,000 model grid points, an observational time length of 24 hours and a localization radius of 80 km. Figure 3.10 shows the estimates over the spatial domain at the final time step (after 1,000 days of simulation), while Figure 3.11 shows the estimates over the last 30 days of the simulation, at location  $x = 20,000$  km. For most variables, there is good agreement between the posterior and prior estimates and the true signal. However, both the posterior and prior state estimates of  $H_s$  and  $P$  fail to capture a few of their peaks. Notice that many of the peaks in  $P$  are slightly underestimated. Indeed, we will see that with satellite observations, these variables are difficult

to estimate.

Next, we examine the sensitivity of the filtering skill with respect to localization radius. In Figure 3.12, we show the spatially and temporally averaged root mean square (RMS) error and the temporally averaged pattern correlation between the posterior (black lines) and prior (gray lines) state estimates of humidity  $q$  and precipitation rate  $P$  and true signal as a function of localization radius, for  $D_{obs} = 40$  km, 80 km and 160 km. In each case, the analysis estimates are more accurate than the background estimates, as we expect, and we obtain the best results with a longer localization radius. For  $D_{obs} = 40$  km (left panels), the improvement after increasing the radius above 80 km (middle panels) is small for each variable. Similarly, for  $D_{obs} = 80$  km, the improvement in RMS error and correlation for both variables is small with increasing radius.

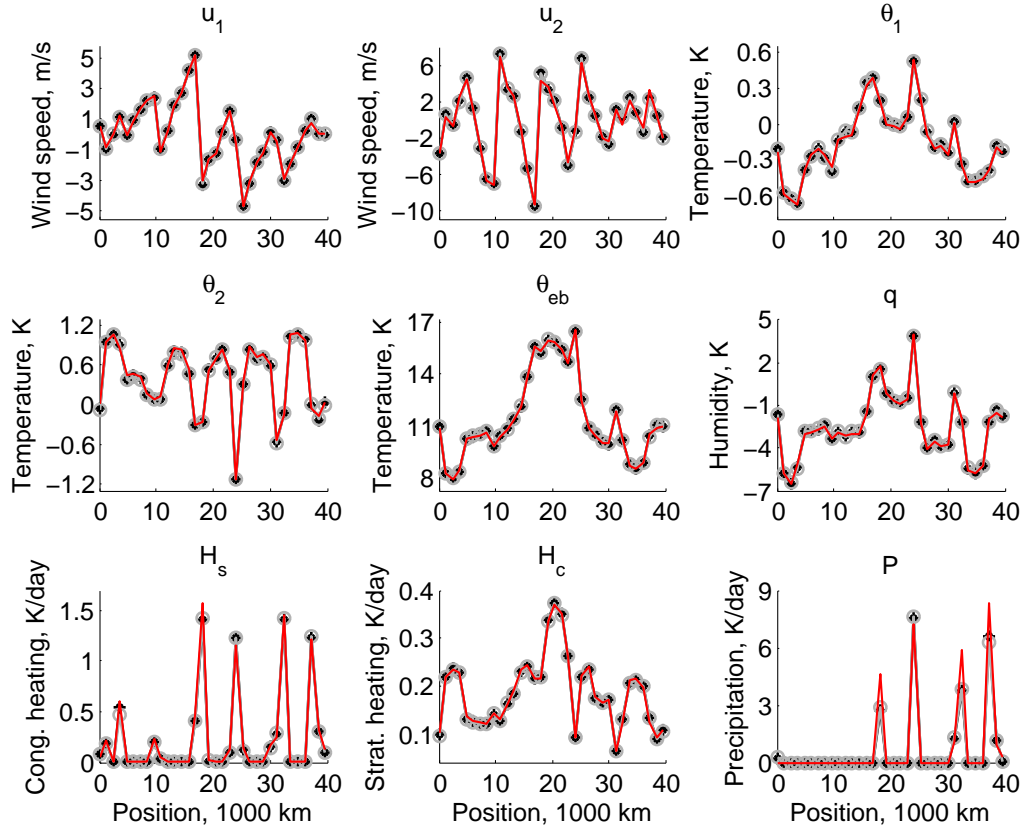


Figure 3.10: Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at time  $T = 1,000$  days, with direct observations of temperature and humidity,  $D_{obs} = 40$  km,  $T_{obs} = 24$  hours,  $K = 16$ , covariance inflation  $r = 19\%$  and localization radius 80 km.

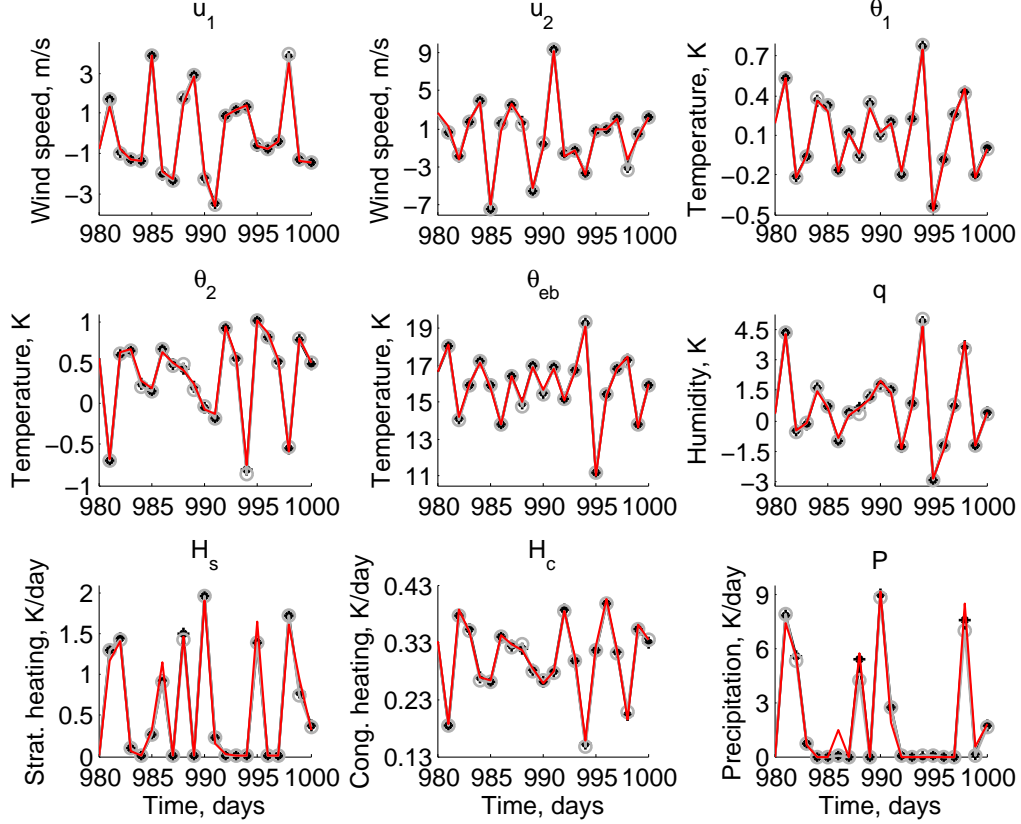


Figure 3.11: Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at location  $x = 20,000$  km over the last 30 days, with direct observations of temperature and humidity,  $D_{obs} = 40$  km,  $T_{obs} = 24$  hours,  $K = 16$ , covariance inflation  $r = 19\%$  and localization radius 80 km.

However, with  $D_{obs} = 160$  km (right panels), there is an obvious improvement in the estimate of  $P$  with increasing radius; its posterior state estimate pattern correlation increases from approximately 0.93 to 0.98 as the radius increases from 0 to 120 km. The RMS errors associated with  $q$  clearly decrease with increasing radius, however, the pattern correlation remains above 0.99, even for a radius of 0.

### 3.5.2 Synthetic Clear Sky AIRS Observations

We now report the numerical results after assimilating synthetic, clear sky AIRS observations. In these experiments, we use two uncorrelated AIRS channels corresponding to the wavenumbers  $1377.4 \text{ cm}^{-1}$  and  $2500.6 \text{ cm}^{-1}$ . For reference, the weighting functions associated with these

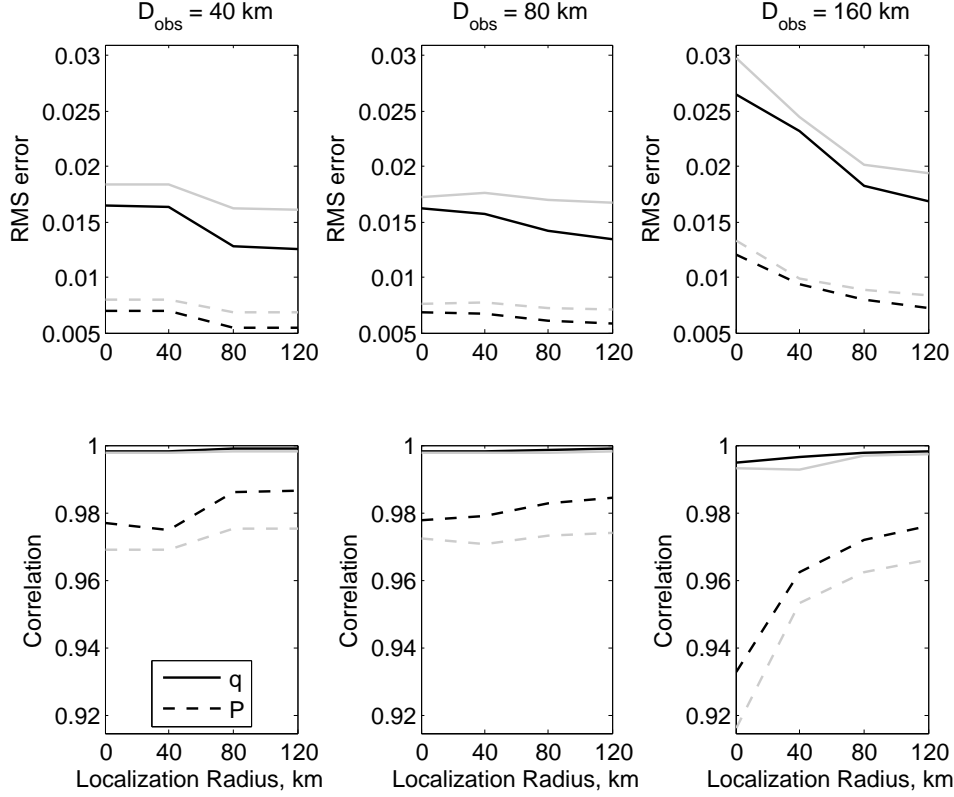


Figure 3.12: Spatially and temporally averaged RMS error and temporally averaged pattern correlation between the posterior and prior state estimates of  $q$  and  $P$  and true signal as a function of localization radius, with direct observations of the temperature and humidity,  $K = 16$ , and  $T_{\text{obs}} = 24$  hours. The black lines correspond to posterior state estimates, while the gray lines correspond to prior state estimates.

channels are given in the right panel of Figure 3.7 and the left panel of Figure 3.9.

We first examine how the filter skill improves with increasing ensemble size. In Figure 3.13, we show the temporally and spatially averaged RMS error between posterior state estimates of the partially observable variable  $q$  and the unobservable variable  $H_c$  and their true signals as a function of ensemble size  $K$ . As  $K$  increases from 8 to 16, there is a clear, albeit slight, decrease in RMS error for each variable. We found that, for every variable, this decrease is small, on the order of  $10^{-2}$  or less, as the ensemble size increases above 16. Therefore, in the remaining experiments, we take  $K = 16$ .

We again show an example of the analysis and background estimates of each variable compared with the true signal, using  $D_{\text{obs}} = 40$  km,  $T_{\text{obs}} = 24$  hours,  $K = 16$  and a localization radius of 80 km. In Figure 3.14, we show the estimates over the spatial domain at the last



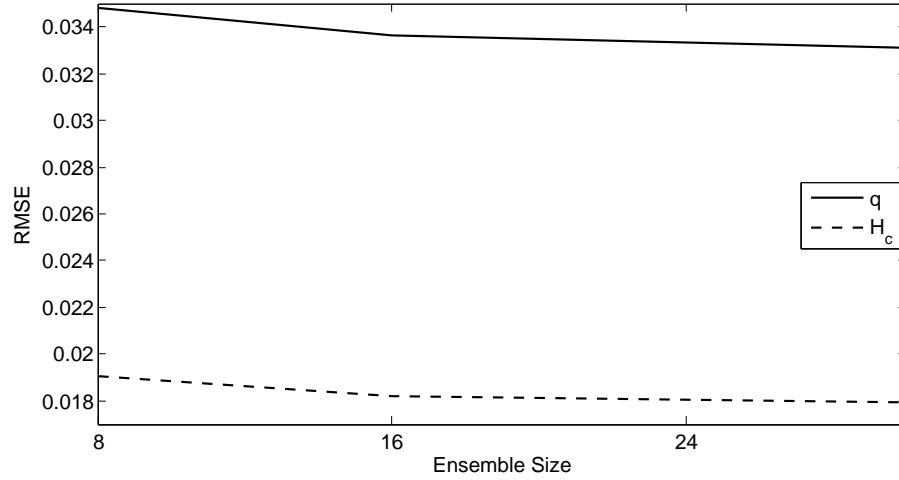


Figure 3.13: RMS errors between the posterior state estimates of  $q$  and  $H_c$  and their true signals as a function of ensemble size, with clear sky observations,  $D_{obs} = 80$  km,  $T_{obs} = 24$  hours, and localization radius 80 km.

time step, and in Figure 3.15, we show the estimates over the last 30 days of the simulation, at the location = 20,000 km. For several variables there again is good agreement between the estimates and the truth. However, for  $H_s$  and  $P$ , which exhibit large peaks, we notice a bigger discrepancy between the truth and the estimates, as compared to the results of assimilating direct observations (see Figures 3.10 and 3.11); The precipitation rate is grossly underestimated at several times. Additionally, in Figure 3.10, the estimates of  $P$  indicate nonzero precipitation when it is actually zero.

The filter estimates are more sensitive to localization radius size when assimilating clear sky AIRS observations as compared to direct observations. Figure 3.16 shows the spatially and temporally averaged RMS error and the temporally averaged pattern correlation between the posterior (black lines) and prior (gray lines) state estimates of  $q$  and  $P$  and their true signals, as a function of localization radius, for  $D_{obs} = 40, 80$ , and 160 km, and  $T_{obs} = 24$  hours. For an observational spacing of 40 km (left panels) and 80 km (middle panels), there is a slight improvement in the estimates of both variables as the radius increases. With a spacing of 40 km the estimates of  $q$  and  $P$  most noticeably improve when the radius increases from 0 to 40 km. However, with  $D_{obs} = 160$  km (right panels), we observe the poorest skill with a radius of 40 km; the RMS error increases (while pattern correlation decreases) as the radius increases from 0 to 40 km. Furthermore, with a radius of 40 km, the prior state estimates of both variables are actually better than the posterior estimates. With such a sparse observation network, it

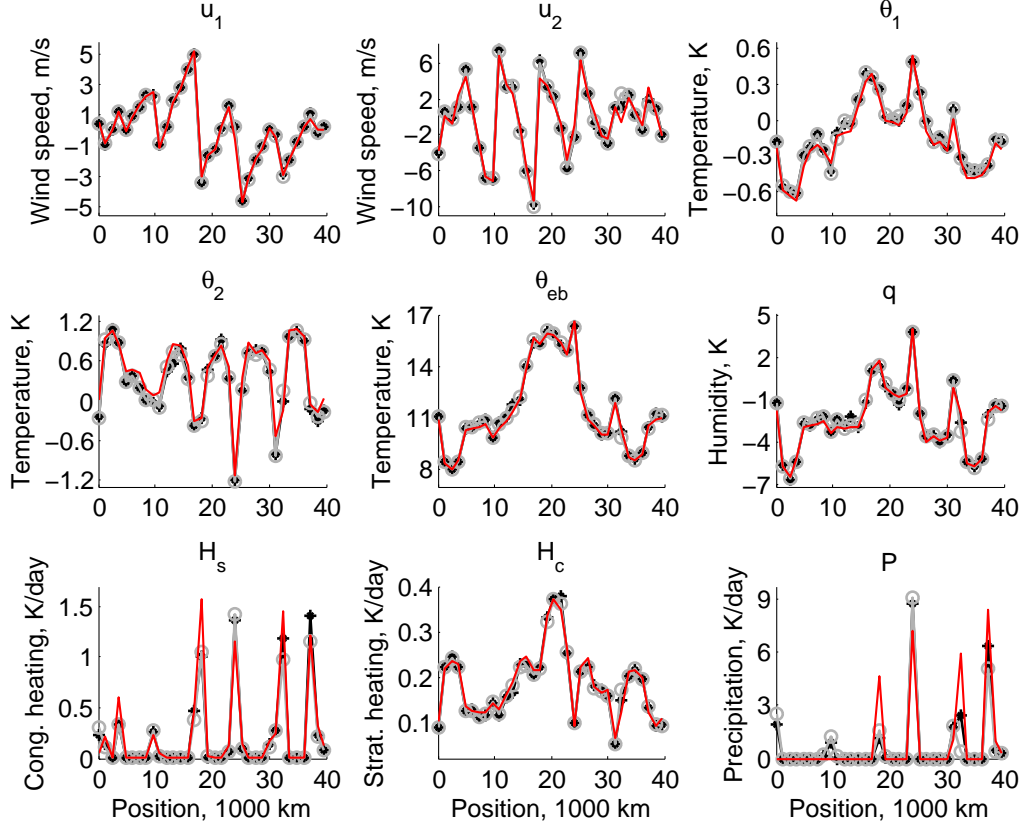


Figure 3.14: Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at time  $T = 1,000$  days, with clear sky observations,  $D_{obs} = 80$  km,  $T_{obs} = 24$  hours,  $K = 16$ , covariance inflation  $r = 16\%$  and localization radius 80 km.

is difficult to achieve a good filter skill. For the observable variable  $q$ , notice that the pattern correlation remains high, above 0.97, in every case except for  $D_{obs} = 160$  km with a radius of 40 km. The pattern correlation associated with  $P$ , however, is lower than it was in the case of direct observations (see Figure 3.12). Here the correlation is consistently below 0.92 and drops as low as 0.7 in one instance.

### 3.5.3 Synthetic Cloudy Sky AIRS Observations

Finally, we present the results achieved after assimilating cloudy sky AIRS observations. To begin, we compare different channel selection schemes; we compare analyses from assimilating the six channels depicted in Figures 3.7-3.9, to the analyses after using two channels at a time, chosen based on the cloud height as described in the previous section. Figure 3.17 shows the

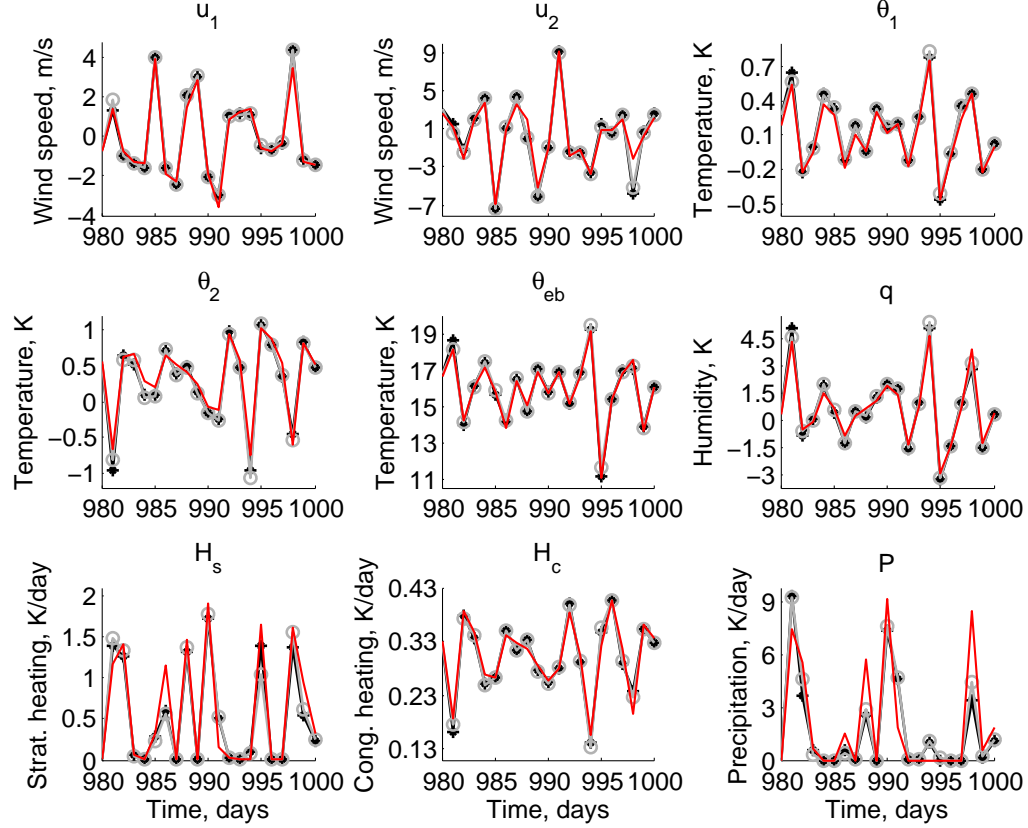


Figure 3.15: Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at location  $x = 20,000$  km over the last 30 days, with clear sky observations,  $D_{obs} = 80$  km,  $T_{obs} = 24$  hours,  $K = 16$ , covariance inflation  $r = 16\%$  and localization radius 80 km.

spatially averaged RMS errors associated with the analysis estimates of  $u_1$ ,  $\theta_2$ ,  $q$  and  $P$  as functions of time, using both channel selection schemes, with perfectly specified cloud cover percent and height. In this simulation, the assimilation is performed with a localization radius of 80 km,  $T_{obs} = 24$  hours and  $D_{obs} = 40$  km. For comparison, we also include the RMS errors based on assimilating these six channels with a clear sky. We observe that in the presence of clouds, using all six channels rather than two cloud-dependent channels improves the filtering skill. Indeed, the RMS error is close to that associated with clear sky observations. This result suggests that if we can ascertain the cloud cover percent and cloud top height, we can improve our filter estimates by assimilating more channels, even those that weight below the cloud tops.

To study the sensitivity of the filtering skill to the error in cloud cover percent and type, we assimilate these six channels after corrupting the true values of  $c$  and  $\Lambda$  with noises of

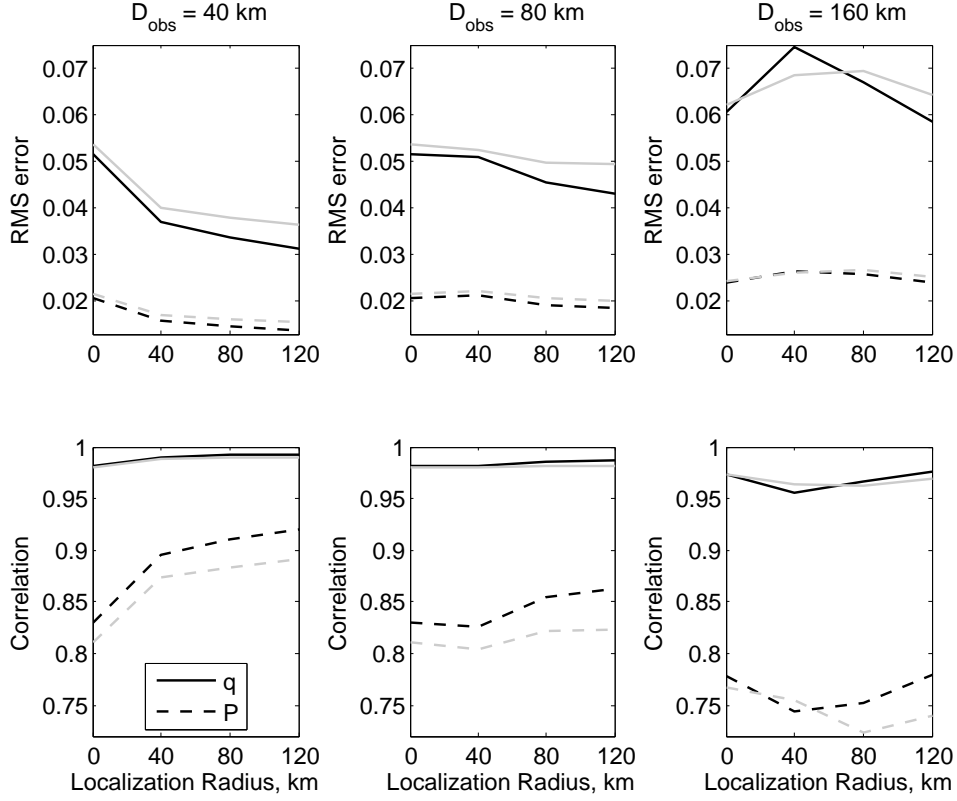


Figure 3.16: Spatially and temporally averaged RMS error and temporally averaged pattern correlation between the posterior (black lines) and prior (gray lines) state estimates of  $q$  and  $P$  and true signal, as a function of localization radius, with  $K = 16$ ,  $T_{\text{obs}} = 24$  and clear sky AIRS observations.

variances equal to ten percent of each variable's respective climatological variance. In Figure 3.18, we show the spatially averaged RMS error associated with  $u_1$ ,  $\theta_2$ ,  $q$  and  $P$  as a function of time, obtained from noisy  $c$ , noisy  $\Lambda$ , and from noisy  $c$  and  $\Lambda$  both. For comparison, we also include the result of assimilating these six channels with perfectly specified  $c$  and  $\Lambda$ . When  $\Lambda$  is corrupted with noise, the cloud type is not accurately specified at times, leading to a large discrepancy in cloud height at these instances. The difference between the observations and the interpolated observations  $h(\xi^b)$  can be large in channels whose weighting functions peak in the lower atmosphere. In this case, the filter skill deteriorates; the RMS error of the posterior estimates of each variable is higher here than when using cloud-dependent channels with the true  $\Lambda$ . In fact, when we use the forecast to approximate  $\Lambda$ , the filter numerically blows up in finite time when using either channel selection scheme. This suggests the cloud height is the

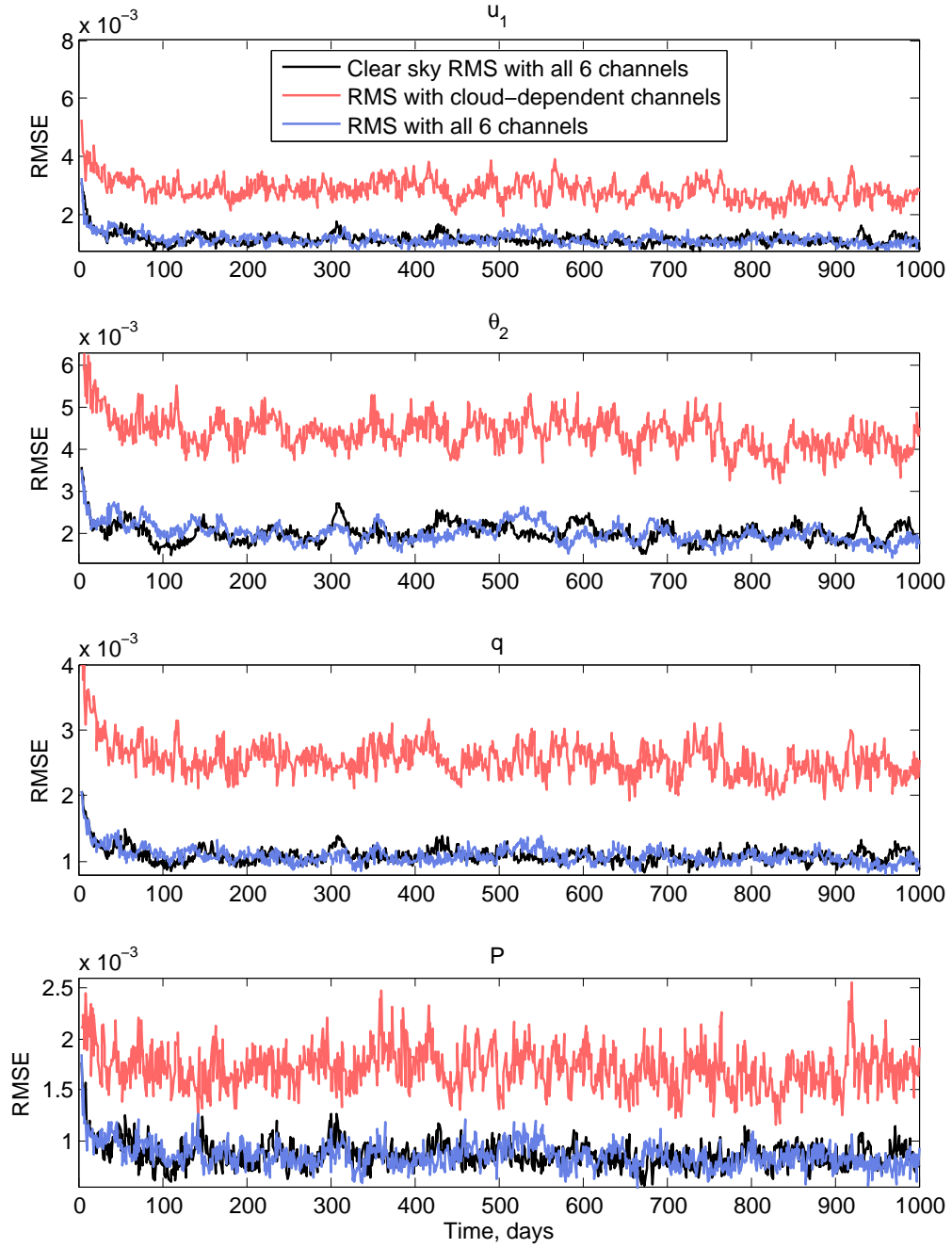


Figure 3.17: Spatially averaged RMSE between the posterior state estimates of  $u_1$ ,  $\theta_2$ ,  $q$  and  $P$  and their true signals over time with  $K = 16$ ,  $T_{obs} = 24$  hours, and localization radius 80 km.

most important parameter to estimate when assimilating cloudy observations.

When the cloud cover percent  $c$  is noisy, but  $\Lambda$  is perfectly specified, the filter skill is comparable to when cloud cover is perfectly specified (compare the red and black curves in Figure 3.18). However, there are a few times when there are large disparities between the two.

If the cloud cover percent cannot be accurately determined, the discrepancy between the interpolated observations,  $h(\xi(x^b))$  in and the observations (the biases) can be large in channels whose weighting functions peak below the cloud top. In our simulations, we find the filter numerically blows up when using the six channels represented by Figures 3.7-3.9 and estimating the cloud cover percent from the background estimates of  $P$  and  $H_c$ , even when the cloud height is perfectly specified. In this case, one can still attain reasonably good estimates by choosing channels that weight only above the cloud. In Figure 3.19, we show the spatially averaged RMS errors associated with the posterior estimates of  $u_1$ ,  $\theta_2$ ,  $q$  and  $P$  as a function of time, after using the background heating rates to estimate the cloud cover percent. We also include the result of assimilating two cloud-dependent channels with perfectly specified  $c$  and  $\Lambda$ , and of assimilating the six channels with a clear field of view. Here the RMS errors of  $P$  and  $u_1$  is as good as when the true cloud cover percent is specified. For the observable variables  $\theta_2$  and  $q$ , the RMS errors are slightly higher when the filter uses its background heating rates to estimate the cloud cover percentage.

Now that we understand how to choose channels for assimilation, we will examine the numerical results obtained by using the channel selection method described in the previous section in more detail. To begin, we compare the posterior and prior state estimates of the multicloud prognostic variables using  $D_{obs} = 40$  km,  $T_{obs} = 24$  hours and a localization radius of 80 km, as we did after assimilating clear sky observations, and direct observations of temperature and humidity. In Figure 3.20 we show the estimates over the spatial domain at the final time step, and in Figure 3.21, we show the estimates over the last 30 days of the simulation, at  $x = 20,000$  km. Compared to the results of assimilating direct observations of temperature and humidity (see Figures 3.10 and 3.11) and of assimilating clear sky AIRS observations (see Figures 3.14 and 3.15), there is a bigger discrepancy between the truth and the analysis and background estimates. The difference between the estimates and the true values of  $P$  and  $H_s$  is more obvious here, especially for  $H_s$ . Here the filter poorly estimates the peaks in  $H_s$  and  $P$ . Also, the filter seems to poorly estimate  $q$  and  $\theta_1$  at times.

To demonstrate that  $H_s$  is the key variable to accurately estimate the precipitation, we show a numerical result in which we insert the true value of  $H_s$  in our data assimilation scheme. This simulation provides a minimum error estimate (in a sense where  $H_s$  is accurately assimilated) for assimilating a mixed set of observations, cloudy AIRS data and direct observations of the stratiform heating rate. In Figure 3.22, we show the posterior and prior state estimates of the other multicloud model variables over the last 30 days of the simulation. Notice here that with

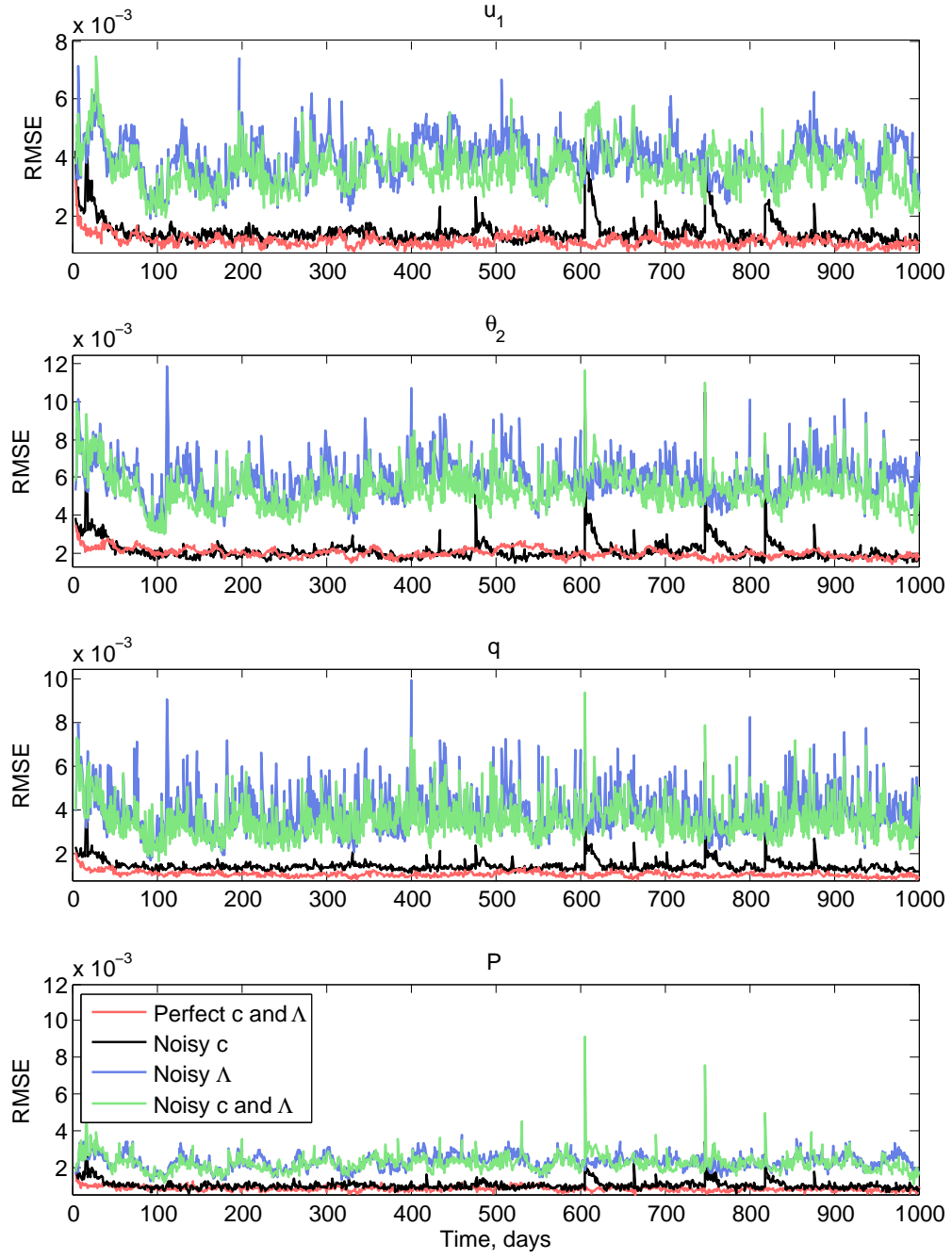


Figure 3.18: Spatially averaged RMSE between the posterior state estimates of  $u_1$ ,  $\theta_2$ ,  $q$  and  $P$  and their true signals over time with  $K = 16$ ,  $T_{obs} = 24$  hours, and localization radius 80 km.

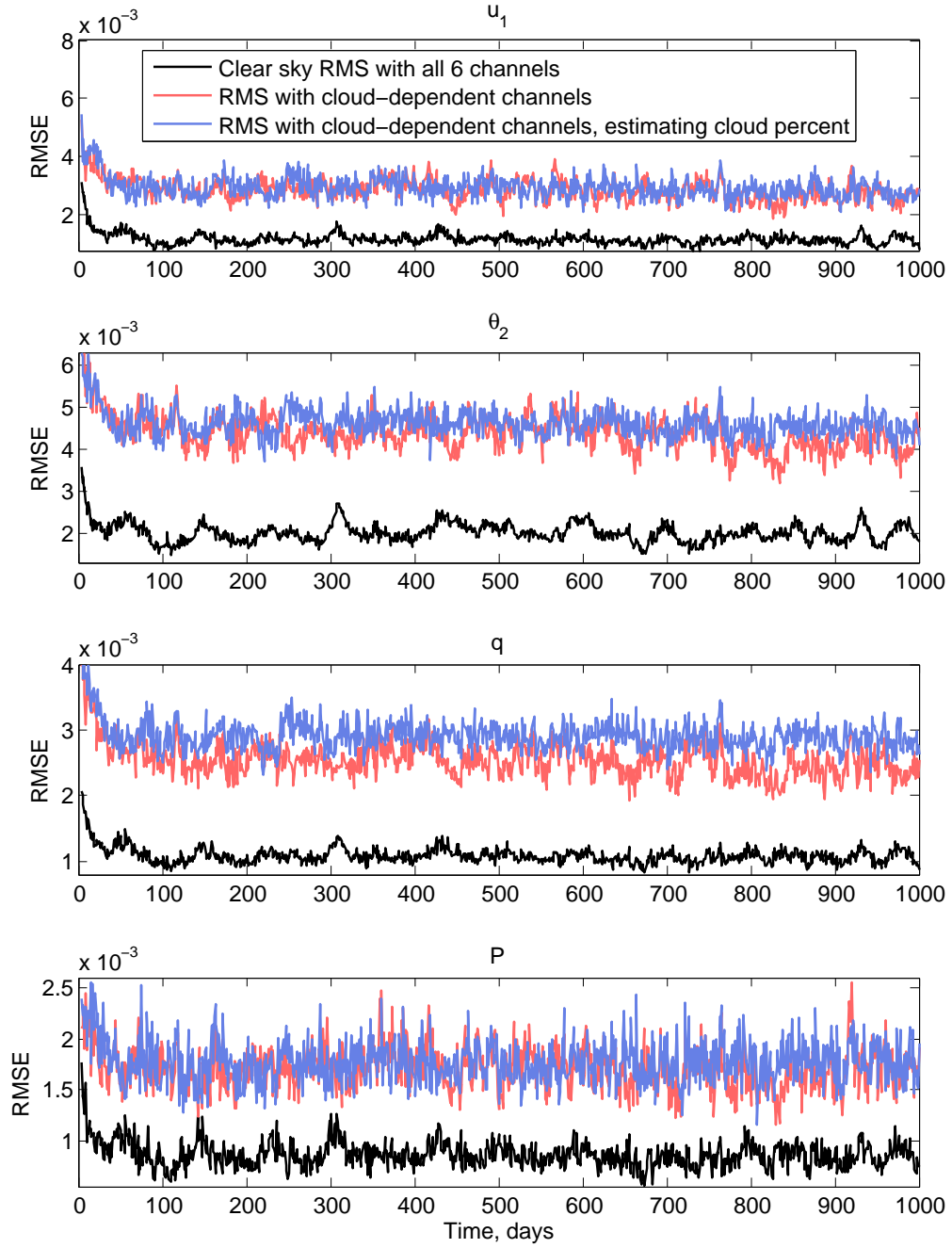


Figure 3.19: Spatially averaged RMSE between posterior state estimates of  $u_1$ ,  $\theta_2$ ,  $q$  and  $P$  and their true signals over time with  $K = 16$ ,  $T_{obs} = 24$  hours, and localization radius 80 km.



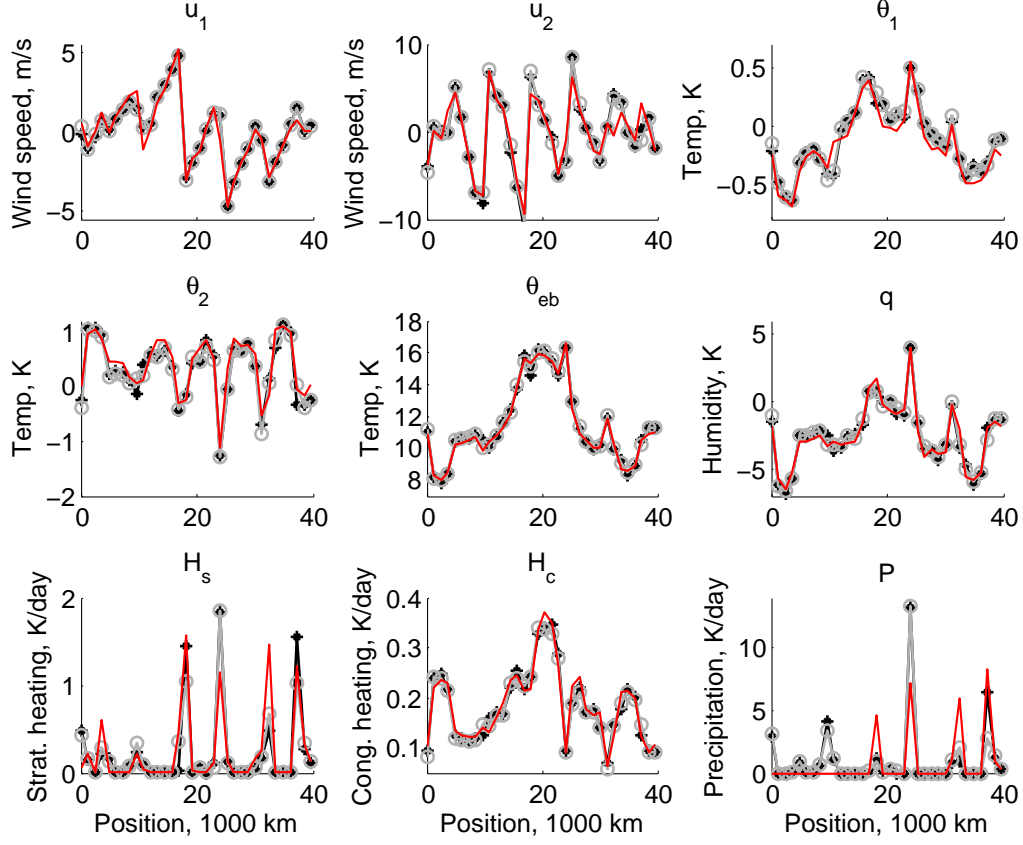


Figure 3.20: Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at time  $T = 1,000$  days, with cloudy sky AIRS observations,  $D_{obs} = 40$  km,  $T_{obs} = 24$  hours,  $K = 16$ , covariance inflation  $r = 14\%$  and localization radius 80 km.

the additional direct observation of  $H_s$ , the filter estimates for  $q$ ,  $\theta_1$  and  $P$  improve. In fact, the filter estimates of all the multicloud prognostic variables have smaller RMS error. We plot the precipitation contours in Figure 3.23, comparing the case in which we assimilate true value of  $H_s$  (bottom right panel) with the case in which we only assimilate noisy AIRS observations (bottom left panel). For reference, we include the true precipitation contours in the top panel. We observe that individual waves are moving toward the warm pool centered at 20,000 km, promoting deep convection at regions within the warm pool. Again, there is a discrepancy between the truth and the estimated precipitation. However, when the true value of  $H_s$  is used in the filter, the estimates improve.

Next we will experiment with different observation networks. In Figure 3.24, we show the

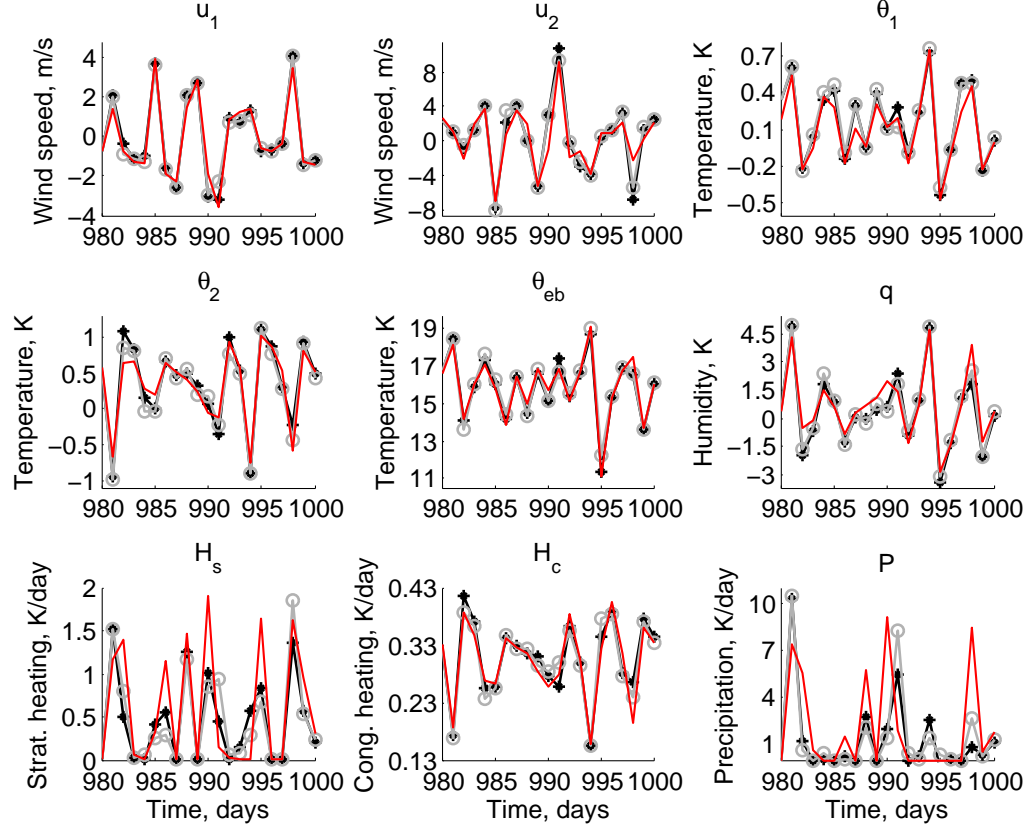


Figure 3.21: Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at location  $x = 20,000$  km over the last 30 days, with cloudy sky AIRS observations,  $D_{obs} = 40$  km,  $T_{obs} = 24$  hours,  $K = 16$ , covariance inflation  $r = 14\%$  and localization radius 80 km.

spatially and temporally averaged RMS error between the posterior (black lines) and prior (gray lines) state estimates of each variable and their true signals as a function of  $D_{obs}$ . For comparison, we include the filter results based on assimilating clear sky observations, and based on assimilating direct observations of temperature and humidity. For each type of observation, and for each variable, there is an increase in RMS error as the observation network becomes sparser. In most cases, the posterior state estimates are more accurate than the prior state estimates, as expected. When the prior state estimate happens to be better than the posterior state estimate, for example, see the RMS error of  $u_1$  with cloudy sky observations in Figure 3.24, the difference is extremely small. Even with a spatially dense clear sky satellite observation network, we cannot attain results that are comparable to using only direct observations. Observe, however, that for most variables, with a denser network, the RMS errors associated

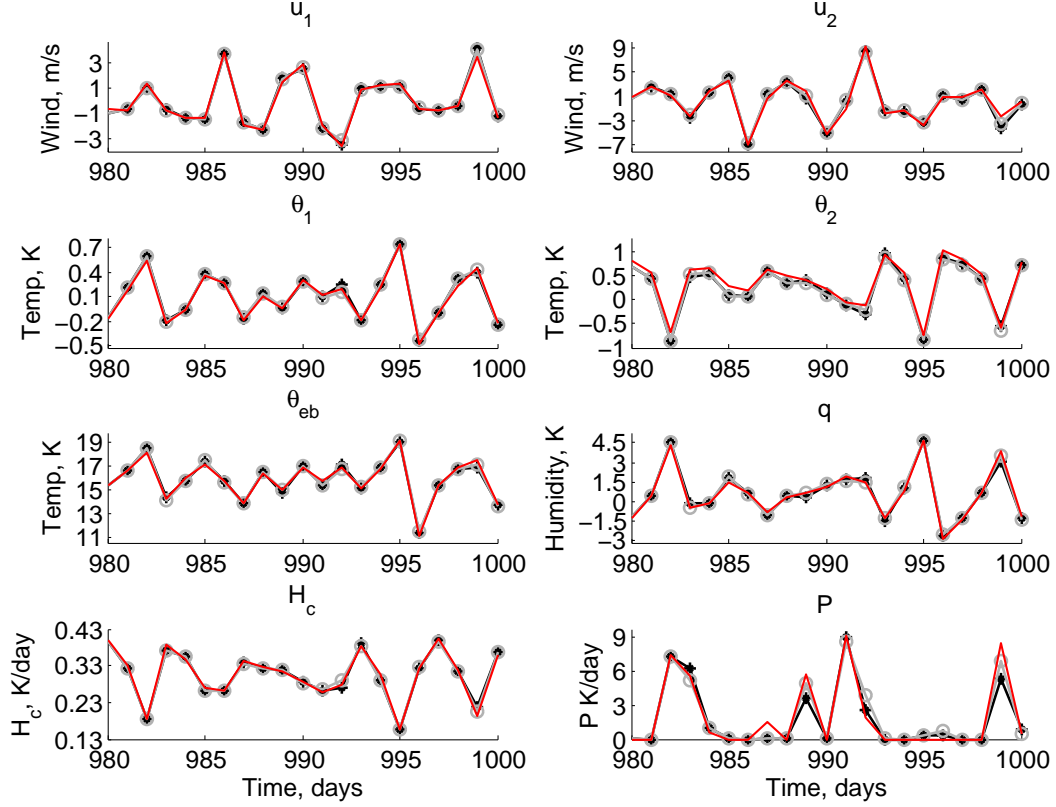


Figure 3.22: Posterior state (black asterisks) and prior state (gray circles) estimates compared with the true signal (red line) at location  $x = 20,000$  km over the last 30 days, with cloudy sky AIRS observations and direct observations of  $H_s$ ,  $D_{obs} = 40$  km,  $T_{obs} = 24$  hours,  $K = 16$  and localization radius 80 km.

with cloudy sky observations are nearly as good as or better than the RMS errors associated with sparser clear sky observations.

In Figure 3.25, we show spatially and temporally averaged RMS error between the analysis (black lines) and background (gray lines) estimates of each multcloud variable and its true signal as a function of  $T_{obs}$ . Again for comparison, we include the results of using clear sky observations, as well as the results of using direct observations of temperature and humidity. For the direct observations, the RMS error associated with each variable, with the exception of  $u_2$  and  $H_c$ , increases as  $T_{obs}$  increases from 6 hours to 12 hours, and then actually decreases as the  $T_{obs}$  increases to 24 hours. In this case, for  $T_{obs} = 6$  hours, the prior state estimates of  $u_1$  and  $u_2$  are more accurate than their respective posterior state estimates. For cloudy sky observations, the RMS error strictly increases with increasing observational time length, and

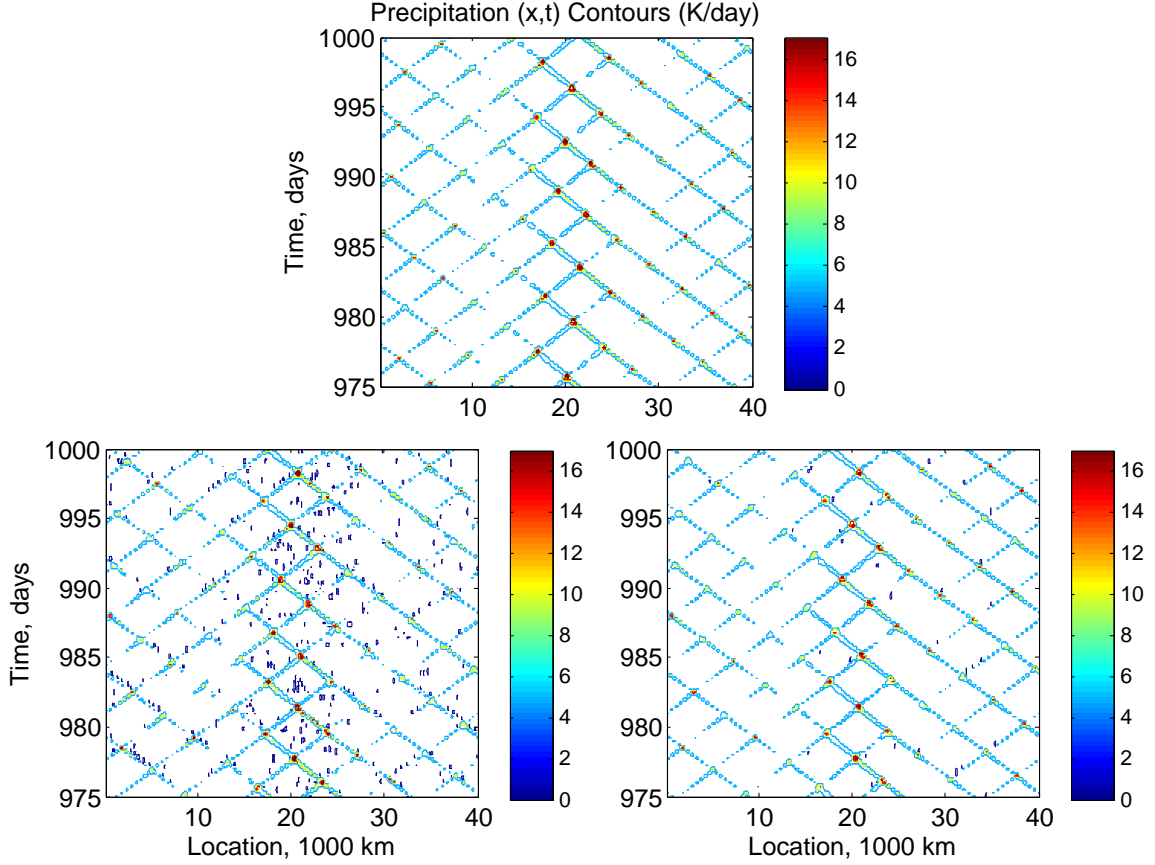


Figure 3.23: The true precipitation contours (top), and estimated precipitation contours (bottom). The bottom right panel shows the result of using the true value of  $H_s$  in the filter. Here,  $T_{obs} = 6$  hours,  $D_{obs} = 40$  km and  $\mathcal{D} = 2$ .

the posterior state estimates are more accurate than prior state estimates, with the exception of  $u_1$ . Notice that for clear sky observations, the increase in RMS error for each variable after increasing  $T_{obs}$  from 12 hours to 24 hours is very small (in the case of  $H_c$ , it actually decreases slightly). To the contrary, the RMS errors associated with cloudy sky observations exhibit a large increase with increasing observational time length. For  $T_{obs} = 6$  hours, the cloudy sky estimates are comparable to the clear sky estimates with  $T_{obs} = 12$  hours. This, together with the findings from Figure 3.24, suggests that assimilating both a temporally and/or spatially denser cloudy sky observation network can produce posterior and prior state estimates that are comparable to estimates obtained from assimilating a temporally and/or spatially sparse clear sky observation network.

The filter estimates with cloudy sky observations do not appear to be sensitive with respect

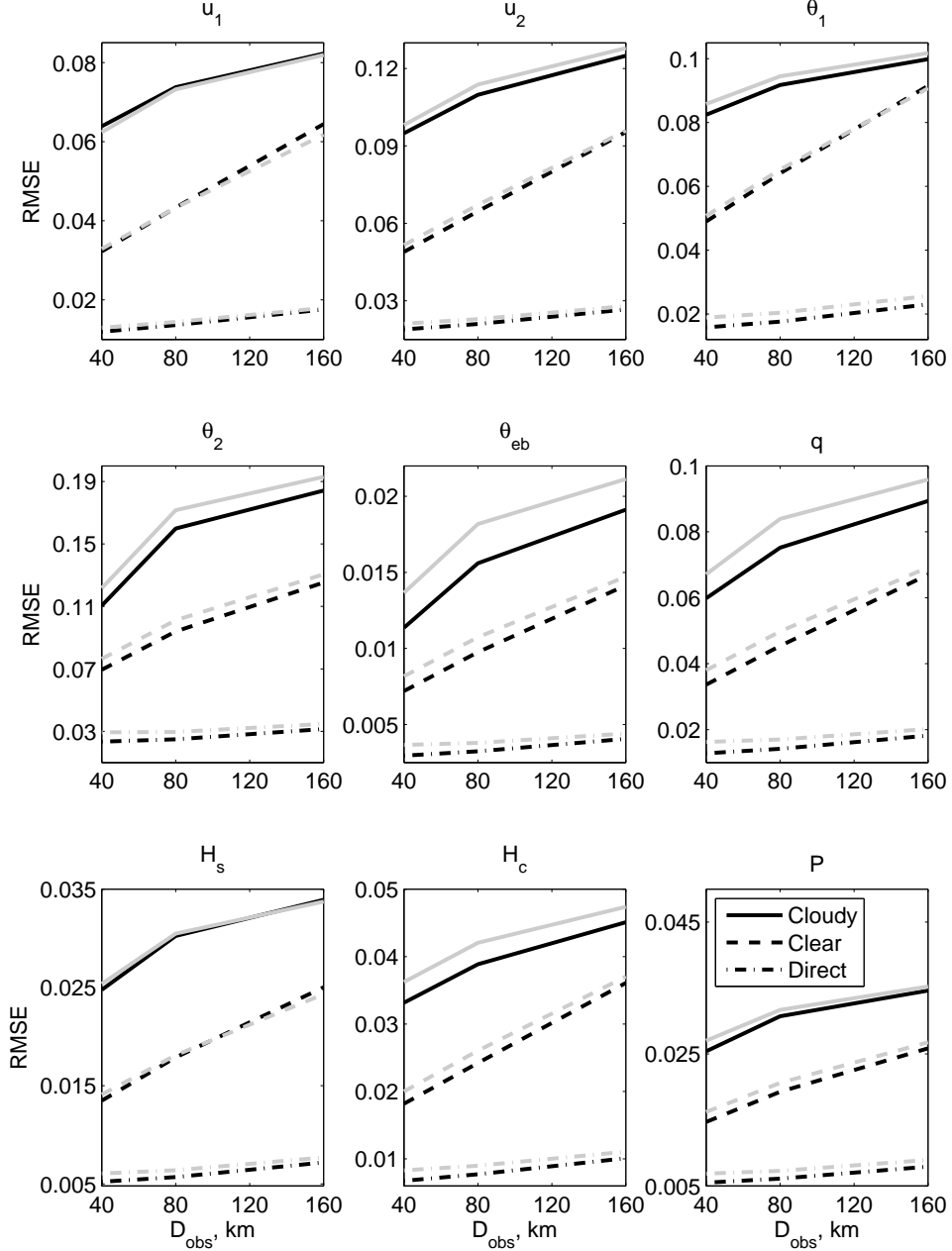


Figure 3.24: Spatially and temporally averaged RMS error between the posterior (black lines) and prior (gray lines) state estimates of the multicloud model prognostic variables and true signal as a function of  $D_{obs}$ , with  $K = 16$ ,  $T_{obs} = 24$  hours, and localization radius 80 km.

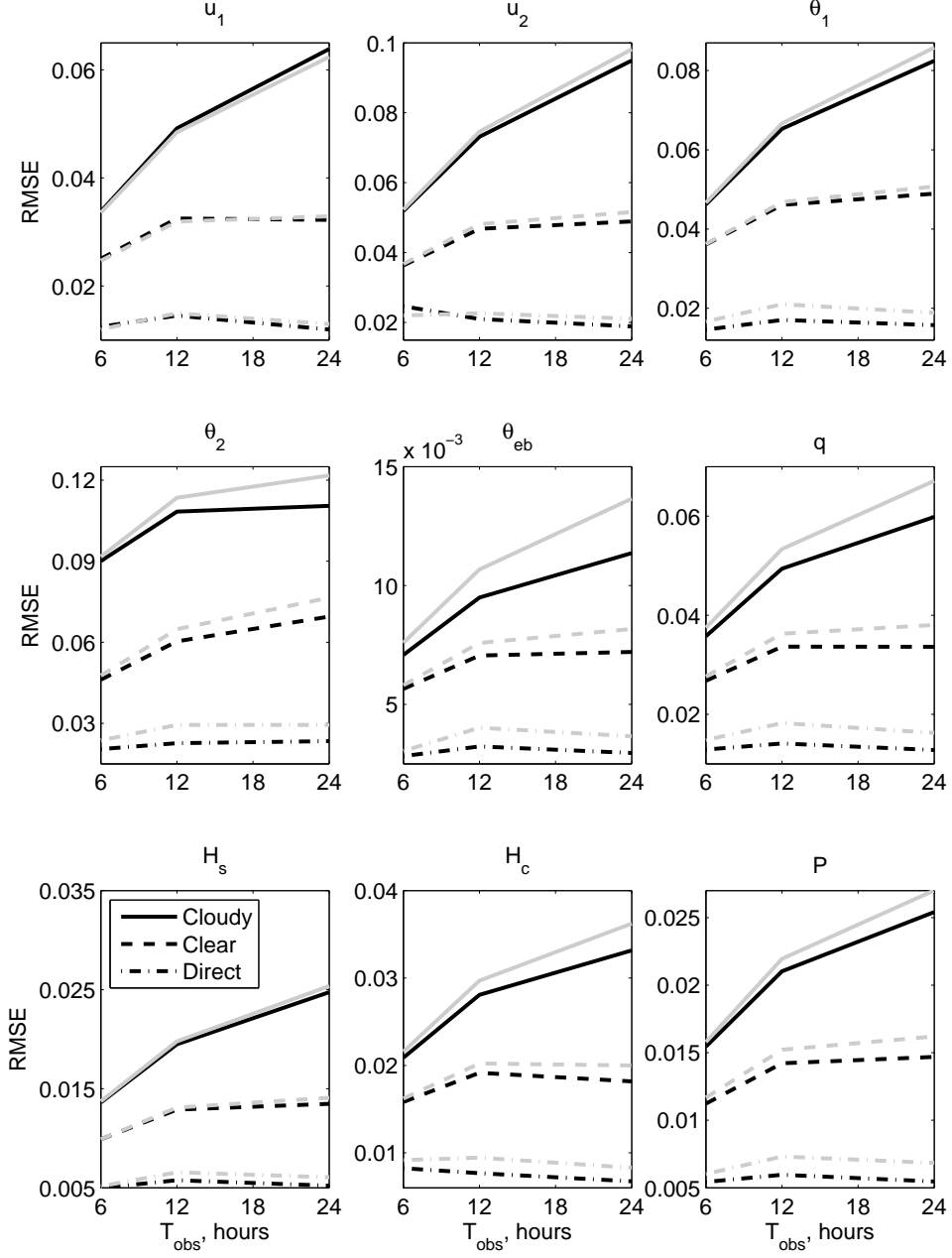


Figure 3.25: Spatially and temporally averaged RMS error between the posterior (black lines) and prior (gray lines) state estimates and true signal as a function of  $D_{obs}$ , with  $K = 16$ ,  $T_{obs} = 24$  hours, and localization radius 80 km.

to localization radius (see Figure 3.26, which shows the average RMS and correlation of  $q$  and  $P$  as a function of localization radius). With a dense observation network, the filter skill improves slightly with a longer radius, while for  $D_{obs} = 80$  km, there is not much difference in filter skill as the radius increases. For a sparse observation network, we appear to have the best filter skill with a localization radius of zero. In this case, for a radius of 40 km, the prior state estimates of both variables are better than the posterior state estimate, which we also observed when using clear sky observations (see the right panels of Figure 3.16).

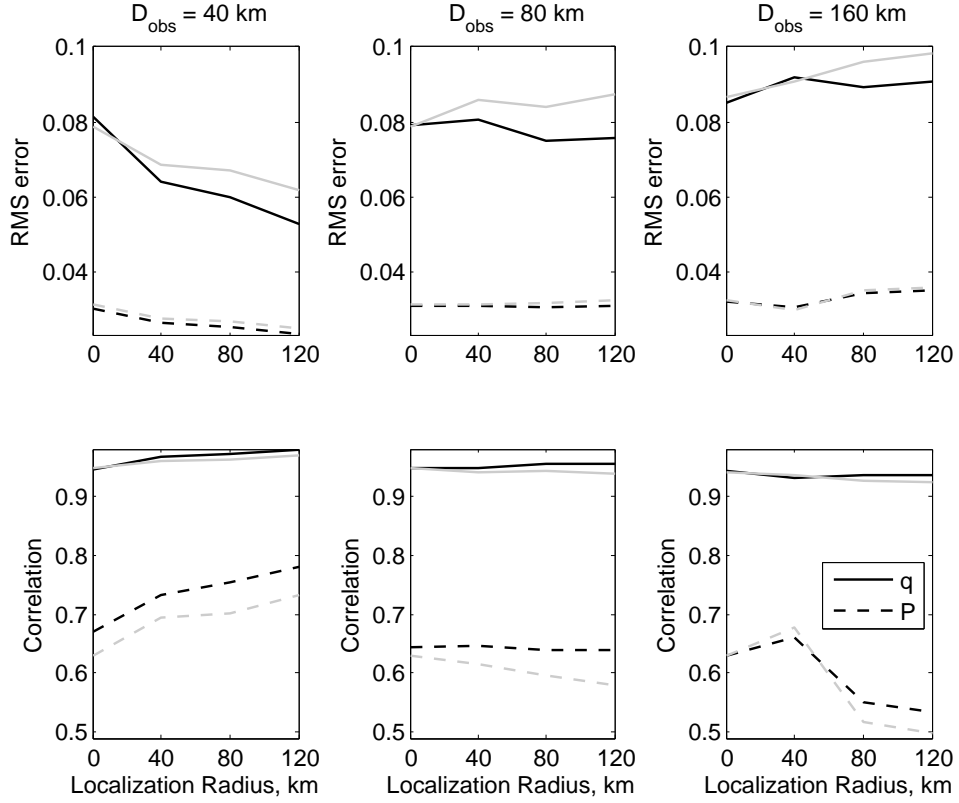


Figure 3.26: Temporally and Spatially averaged RMS error and temporally averaged pattern correlation between the posterior (black lines) and prior (gray lines) state estimates and true signal as a function of localization radius, with cloudy sky AIRS observations,  $T_{obs} = 24$  hours, and  $K = 16$ .

### 3.6 Summary

In this chapter, we assessed the potential for using cloudy sky AIRS observations to improve practical weather forecasting skill and identified the key parameters that induce biases in the filter estimates in the presence of, at most, one cloud at a time. Our study was based on synthetic cloudy AIRS observations, generated from a numerical solution of the multicloud model, [32, 33], a toy model of convectively coupled tropical waves that incorporates cloud dynamics at the equator. This model has three variables, congestus heating rate,  $H_c$ , stratiform heating rate,  $H_s$  and precipitation rate  $P$ , that respectively correlate to the presence of three cloud types that are prevalent in the tropics. In our experiments, we used these heating rates and the multicloud model switching function  $\Lambda$  to simulate cloud cover in the satellite data. The cloudy AIRS brightness temperatures were calculated by inserting the true values of temperature, moisture and heating rates into the prototype Community Radiative Transfer Model. Using these cloudy observations, we assessed the filter estimates in different scenarios, incorporating either the true cloud cover percent  $c$  and  $\Lambda$ , noisy  $c$  and/or noisy  $\Lambda$ , or an estimate of  $c$ , calculated from the background estimates.

From our numerical study, we conclude that the most important parameter to estimate in assimilating cloudy AIRS data is the cloud top pressure (which correlates with the cloud type in our experiments). Without a decent estimate of  $\Lambda$ , the biases become too large and can lead to filter divergence (solutions which blow up in finite time). The second most important parameter is the cloud cover percent. If this parameter can be reasonably estimated in addition to the cloud height, we can use more channels, including those with weighting functions that peak below the cloud. However, if the cloud percent cannot be ascertained, one can still obtain reasonable results by assimilating channels whose weighting function peak in the upper atmosphere and are negligible below the cloud top, since these channels are less sensitive to cloud effects [54]. In our numerical experiments, the filter diverges in finite time when we assimilate channels with weighting functions that peak below the clouds, when the cloud cover percents are estimated from the filter background estimates.

We also found that if we only assimilate AIRS channels whose weighting functions peak in the upper part of the atmosphere, the state estimation of the precipitation rate is not completely recovered, even when the cloud type and coverage percent are perfectly specified. We demonstrated that additional direct observations of the prognostic variables can improve the filter estimates; there was a dramatic improvement in the precipitation estimation when direct observations of the true stratiform heating rate were assimilated in addition to cloudy AIRS data. When using only direct observations of temperature and humidity, the estimates of each variable, including precipitation, are quite accurate. We conclude that different types of observations, especially direct observations of prognostic variables, can help uncover the observability



issues that arise in assimilating cloudy sky AIRS data.

Assimilating a temporally (or spatially) denser network of cloudy observations produces analysis and background estimates that are comparable to the estimates obtained after assimilating a temporally (or spatially) sparse clear sky observation network. When the observations are spatially dense, we find that the filter estimates improve as the localization radius increases. In contrast, for a spatially sparse network, the filter estimates degrade as the localization radius increases. This is true for each type of observation considered in this chapter: direct observations and both clear and cloudy sky AIRS radiances.

Currently, the only AIRS observations that are used in weather systems are the cloud free channels [60]. Although AIRS data are believed to be important for improving weather prediction, the percentage of AIRS observations that are assimilated remains small. The findings here can provide useful guidelines as well as the limitations for using cloud AIRS observations in practical data assimilation. Our encouraging results obtained by using channels that observe the entire atmospheric column rely on accurate estimates of the cloud cover percent and cloud top pressure. In reality, cloud top pressure can be accurately estimated by CALIPSO [64], while the estimation of cloud cover percent is challenging. Of course, biases can also arise from other issues that are not discussed here, such as the satellite scan angle, calibration error, and modeling errors in the radiative transfer model. The issue of bias correction is addressed in [5, 57, 52, 16]. There are also other parameters that are not considered here, such as cloud mask [59].

## REFERENCES

- [1] R. S. Ajayamohan, B. Khouider, and A.J. Majda. Realistic initiation and dynamics of the madden-julian oscillation in a coarse resolution aquaplanet gcm. *Geophysical Research Letters*, 40(23):6252–6257, 2013.
- [2] Jeffrey L Anderson. An ensemble adjustment kalman filter for data assimilation. *Monthly weather review*, 129(12), 2001.
- [3] Jeffrey L Anderson and Stephen L Anderson. A monte carlo implementation of the non-linear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12), 1999.
- [4] Joséé A Aravéequia, Istvan Szunyogh, Elana J Fertig, Eugenia Kalnay, David Kuhl, and Eric J Kostelich. Evaluation of a strategy for the assimilation of satellite radiance observations with the local ensemble transform kalman filter. *Monthly Weather Review*, 139(6), 2011.
- [5] T Auligné, AP McNally, and DP Dee. Adaptive bias correction for satellite data in a numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, 133(624):631–642, 2007.
- [6] S.-J. Baek, B.R. Hunt, E. Kalnay, E. Ott, and I. Szunyogh. Local ensemble Kalman filtering in the presence of model bias. *Tellus A*, 58(3):293–306, 2006.
- [7] Craig H Bishop, Brian J Etherton, and Sharanya J Majumdar. Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly weather review*, 129(3):420–436, 2001.
- [8] K Brown and J Harlim. Assimilating irregularly spaced sparsely observed turbulent signals with hierarchical bayesian reduced stochastic filters. *Journal of Computational Physics*, 235:143–160, 2013.
- [9] E Castronovo, J Harlim, and AJ Majda. Mathematical test criteria for filtering complex systems: plentiful observations. *Journal of Computational Physics*, 227(7):3678–3714, 2008.
- [10] N. Cressie and C.K. Wikle. *Statistics for spatio-temporal data*. Wiley, New Jersey, 2011.
- [11] Timothy DelSole. Stochastic models of quasigeostrophic turbulence. *Surveys in Geophysics*, 25(2):107–149, 2004.
- [12] T.J. Dunkerton and F.X. Crum. Eastward propagating 2- to 15-day equatorial convection and its relation to the tropical intraseasonal oscillation. *J. Geophys. Res.*, 100:25781–25790, 1995.
- [13] Moustafa T. Chahine et al. AIRS improving weather forecasting and providing new data on greenhouse gases. *Bulletin of the American Meteorological Society*, 87(7):911–926, 2006.

- [14] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans (1978–2012)*, 99(C5):10143–10162, 1994.
- [15] J. R. Eyre and P. D. Watts. A sequential estimation approach to cloud-clearing for satellite temperature sounding. *Quarterly Journal of the Royal Meteorological Society*, 113(478):1349–1376, 1987.
- [16] Elana J Fertig, SEUNG-JONG BAEK, Brian R Hunt, Edward Ott, Istvan Szunyogh, José A Aravéquia, Eugenia Kalnay, Hong Li, and Junjie Liu. Observation bias correction with an ensemble kalman filter. *Tellus A*, 61(2):210–226, 2009.
- [17] Elana J Fertig, Brian R Hunt, Edward Ott, and Istvan Szunyogh. Assimilating non-local observations with a local ensemble kalman filter. *Tellus A*, 59(5):719–730, 2007.
- [18] D.M. Frierson, A.J. Majda, and O.M. Pauluis. Large scale dynamics of precipitation fronts in the tropical atmosphere: A novel relaxation limit. *Comm. Math. Sci.*, 2(4):591–626, 2004.
- [19] E. Fuchs and D. J. Raymond. Large-scale modes of a nonrotating atmosphere with water vapor and cloud-radiation feedbacks. *J. Atmos. Sci.*, 59:1669–1679, May 2002.
- [20] Louis Garand and Serge Nadon. High-resolution satellite analysis and model evaluation of clouds and radiation over the mackenzie basin using avhrr data. *Journal of climate*, 11(8), 1998.
- [21] P.T. Haertel and G.N. Kiladis. On the dynamics of two-day equatorial disturbances. *J. Atmos. Sci.*, 61:2707–2721, 2004.
- [22] Yong Han, Paul van Delst, Quanhua Liu, Fuzhong Weng, Banghua Yan, and John Derber. Users guide to the jcsda community radiative transfer model (beta version). *Joint Center for Satellite Data Assimilation, Camp Springs, MD, USA*, 2005.
- [23] J. Harlim. Interpolating irregularly spaced observations for filtering turbulent complex systems. *SIAM J. Sci. Comp.*, 33(5):2620–2640, 2011.
- [24] J. Harlim and A.J. Majda. Mathematical strategies for filtering complex systems: Regularly spaced sparse observations. *Journal of Computational Physics*, 227(10):5304–5341, 2008.
- [25] J. Harlim and A.J. Majda. Filtering turbulent sparsely observed geophysical flows. *Monthly Weather Review*, 138(4):1050–1083, 2010.
- [26] J. Harlim and A.J. Majda. Test models for filtering and prediction of moisture-coupled tropical waves. *Q. J. Roy. Met. Soc.*, 139:119–136, 2013.
- [27] Sylvain Heilliette, Yves J Rochon, Louis Garand, and Jacek W Kaminski. Assimilation of infrared radiances in the context of observing system simulation experiments. *Journal of Applied Meteorology & Climatology*, 52(4), 2013.
- [28] B.R. Hunt, E.J. Kostelich, and I. Szunyogh. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D*, 230:112–126, 2007.

- [29] R.H. Johnson, Rickenbach T, M, S.A. Rutledge, P.E. Ciesielski, and W.H. Schubert. Tri-modal characteristics of tropical convection. *J. Climate*, 12:2397–2407, 1999.
- [30] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*. Springer, New York, 2005.
- [31] B. Khouider and A. J. Majda. Multicloud convective parameterizations with crude vertical structure. *Theor. Comp. Fluid Dyn.*, 20:351–375, 2006.
- [32] B. Khouider and A.J. Majda. A simple multicloud parametrization for convectively coupled tropical waves. Part I: Linear Analysis. *Journal of the Atmospheric Sciences*, 63:1308–1323, 2006.
- [33] B. Khouider and A.J. Majda. A simple multicloud parametrization for convectively coupled tropical waves. Part II: Nonlinear Simulations. *Journal of the Atmospheric Sciences*, 64:381–400, 2007.
- [34] Boualem Khouider, Andrew J Majda, and Samuel N Stechmann. Climate science in the tropics: waves, vortices and pdes. *Nonlinearity*, 26(1):R1, 2013.
- [35] G.N. Kiladis, K.H. Straub, and P. Haertel. Zonal and vertical structure of the Madden-Julian oscillation. *J. Atmos. Sci.*, 62:2790–2809, 2005.
- [36] R. Kleeman and A.J. Majda. Predictability in a model of geophysical turbulence. *Journal of the Atmospheric Sciences*, 62:2864–2879, 2005.
- [37] Kody JH Law and AM Stuart. Evaluating data assimilation algorithms. *Monthly Weather Review*, 140(11), 2012.
- [38] J Le Marshall, J Jung, J Derber, M Chahine, R Treadon, SJ Lord, M Goldberg, W Wolf, HC Liu, J Joiner, et al. Improving global analysis and forecasting with AIRS. *Bulletin of the American Meteorological Society*, 87(7):891–894, 2006.
- [39] J Le Marshall, J Jung, T Zapotocny, J Derber, R Treadon, S Lord, M Goldberg, and W Wolf. The application of airs radiances in numerical weather prediction. *Australian Meteorological Magazine*, 55(3), 2006.
- [40] J. Li, Chian-Yi Liu, H.-L. Huang, T.J. Schmit, Xuebao Wu, W.P. Menzel, and J.J. Gurka. Optimal cloud-clearing for airs radiances using modis. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(6):1266–1278, June 2005.
- [41] X. Lin and R.H. Johnson. Kinematic and thermodynamic characteristics of the flow over the western Pacific Warm Pool during TOGA COARE. *J. Atmos. Sci.*, 53:695–715, 1996.
- [42] K. N. Liou. *An introduction to atmospheric radiation*, volume 84 of *International Geophysics Series*. Academic Press, 2002.
- [43] A Majda and I Timofeyev. Low-dimensional chaotic dynamics versus intrinsic stochastic noise: a paradigm model. *Physica D: Nonlinear Phenomena*, 199(3):339–368, 2004.

- [44] A. J. Majda and M. G. Shefter. Waves and instabilities for model tropical convective parameterizations. *J. Atmos. Sci.*, 58:896–914, 2001.
- [45] A.J. Majda and J. Harlim. *Filtering Complex Turbulent Systems*. Cambridge University Press, UK, 2012.
- [46] A.J. Majda, S.N. Stechmann, and B. Khouider. Madden-Julian Oscillation analog and intraseasonal variability in a multcloud model above the equator. *Proceedings of the National Academy of Sciences*, 104(24):9919–9924, 2007.
- [47] Andrew J Majda, Christian Franzke, and Boualem Khouider. An applied mathematics perspective on stochastic modelling for climate. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366(1875):2427–2453, 2008.
- [48] Andrew J Majda, Boris Gershgorin, and Yuan Yuan. Low-frequency climate response and fluctuation–dissipation theorems: Theory and practice. *Journal of the Atmospheric Sciences*, 67(4), 2010.
- [49] Andrew J Majda and Michael G Shefter. Models for stratiform instability and convectively coupled waves. *Journal of the atmospheric sciences*, 58(12):1567–1584, 2001.
- [50] Andrew J Majda, Ilya Timofeyev, and Eric Vanden Eijnden. Models for stochastic climate prediction. *Proceedings of the National Academy of Sciences*, 96(26):14687–14691, 1999.
- [51] Andrew J Majda, Ilya Timofeyev, and Eric Vanden-Eijnden. Systematic strategies for stochastic mode reduction in climate. *Journal of the Atmospheric Sciences*, 60(14), 2003.
- [52] AP McNally. The direct assimilation of cloud-affected satellite infrared radiances in the ecmwf 4d-var. *Quarterly Journal of the Royal Meteorological Society*, 135(642):1214–1229, 2009.
- [53] AP McNally and PD Watts. A cloud detection algorithm for high-spectral-resolution infrared sounders. *Quarterly Journal of the Royal Meteorological Society*, 129(595):3411–3423, 2003.
- [54] AP McNally, PD Watts, J A Smith, R Engelen, GA Kelly, JN Thépaut, and M Matri-cardi. The assimilation of airs radiance data at ecmwf. *Quarterly Journal of the Royal Meteorological Society*, 132(616):935–957, 2006.
- [55] J.D. Neelin and N. Zeng. A quasi-equilibrium tropical circulation model-formulation. *Journal of the Atmospheric Sciences*, 57:1741–1766, 2000.
- [56] Jason A Otkin. Clear and cloudy sky infrared brightness temperature assimilation using an ensemble kalman filter. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 115(D19), 2010.
- [57] EG Pavelin, SJ English, and JR Eyre. The assimilation of cloud-affected infrared satellite radiances for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 134(632):737–749, 2008.

- [58] M. Pidwirny. *The Layered Atmosphere, Fundamentals of Physical Geography*. Second edition, 2006.
- [59] Rosanne Polkinghorne and Tomislava Vukicevic. Data assimilation of cloud-affected radiances in a cloud-resolving model. *Monthly Weather Review*, 139(3), 2011.
- [60] O. Reale, J. Susskind, R. Rosenberg, E. Brin, E. Liu, L. P. Riishojgaard, J. Terry, and J. C. Jusem. Improving forecast skill by assimilation of quality-controlled airs temperature retrievals under partially cloudy conditions. *Geophys. Res. Lett.*, 35(8), 2008.
- [61] R. Salmon. *Lectures on geophysical fluid dynamics*, volume 378. Oxford University Press, 1998.
- [62] S.K. Smith, G. Boccaletti, C.C. Henning, I.N. Marinov, C.Y. Tam, I.M. Held, and G.K. Vallis. Turbulent diffusion in the geostrophic inverse cascade. *Journal of Fluid Mechanics*, 469:13–48, 2002.
- [63] K.H. Straub and G.N. Kiladis. Observations of a convectively-coupled Kelvin waves in the eastern pacific ITCZ. *J. Atmos. Sci.*, 59:30–53, 2002.
- [64] CJ Stubenrauch, WB Rossow, S Kinne, S Ackerman, G Cesana, H Chepfer, L Di Girolamo, B Getzewich, A Guignard, A Heidinger, et al. Assessment of global cloud datasets from satellites. *Bulletin of the American Meteorological Society*, 94(7), 2013.
- [65] G.K. Vallis. *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*. Cambridge University Press, Cambridge, U.K., 2006.
- [66] Xuguang Wang and Craig H Bishop. A comparison of breeding and ensemble transform kalman filter ensemble forecast schemes. *Journal of the atmospheric sciences*, 60(9), 2003.
- [67] M. Wheeler and G.N. Kiladis. Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber-frequency domain. *Journal of the Atmospheric Sciences*, 56:374–399, 1999.
- [68] Jeffrey S Whitaker and Thomas M Hamill. Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7), 2002.
- [69] C. Zhang. Madden–Julian Oscillation. *Reviews of Geophysics*, 43:G2003+, June 2005.