

# Test models for filtering and prediction of moisture-coupled tropical waves

John Harlim<sup>a\*</sup> and Andrew J. Majda<sup>b</sup>

<sup>a</sup> Department of Mathematics, North Carolina State University, USA

<sup>b</sup> Department of Mathematics and Center for Atmospheric and Ocean Science, Courant Institute of Mathematical Sciences, New York University, USA

\*Correspondence to: Department of Mathematics, North Carolina State University, BOX 8205, Raleigh, NC 27695, USA. E-mail: jharlim@ncsu.edu

The filtering/data assimilation and prediction of moisture coupled tropical waves is a contemporary topic with significant implications for extended range forecasting. The development of efficient algorithms to capture such waves is limited by the unstable multiscale features of tropical convection which can organize large scale circulations and the sparse observations of the moisture coupled wave in both the horizontal and vertical. The approach proposed here is to address these difficult issues of data assimilation and prediction through a suite of analog models which despite their simplicity capture key features of the observational record and physical processes in moisture coupled tropical waves. The analog models emphasize here involve the multcloud convective parameterization based on three cloud types, congestus, deep, and stratiform, above the boundary layer. Two test examples involving an MJO-like turbulent traveling wave and the initiation of a convectively coupled wave train are introduced to illustrate the approach. A suite of filters with judicious model errors for data assimilation of sparse observations of tropical waves, based on linear stochastic models in a moisture coupled eigenmode basis is developed here and applied to the two test problems. Both the reduced filter and 3D-VAR with a full moist background covariance matrix can recover the unobserved troposphere humidity and precipitation rate; on the other hand, 3D-VAR with a dry background covariance matrix fails to recover these unobserved variables. The skill of the reduced filtering methods in recovering the unobserved congestus and stratiform heating rates as well as the front to rear tilt of the convectively coupled waves exhibits a subtle dependence on the sparse observation network and the observation time. Copyright © 0000 Royal Meteorological Society

*Key Words:* tropical data assimilation, reduced stochastic filters, multcloud models, Madden-Julian Oscillation, convectively coupled waves

*Received ...*

*Citation: ...*

## 1. Introduction

Observational data indicate that through the complex interaction of heating and moist convection, tropical atmosphere flows are organized on a hierarchy of scales (Nakazawa 1988) ranging from cumulus clouds of a few kilometers to mesoscale convective systems (Houze 2004) to equatorial synoptic-scale convectively coupled Kelvin waves and two-day waves (Kiladis *et al.* 2009) to planetary-scale intraseasonal organized circulations such as the Madden-Julian Oscillation (MJO, Zhang 2005). These moisture coupled tropical waves like the MJO exert a substantial influence on intraseasonal prediction in the tropics, sub-tropics, and midlatitudes (Monterey *et al.* 2007). Despite the continued research efforts by the climate community, the present coarse resolution GCM's, used for prediction of weather and climate, poorly represent variability associated with tropical convection (Lau and Waliser 2005; Zhang 2005; Lin *et al.* 2006). Given the importance of moisture coupled tropical waves for short term climate and medium to long range weather prediction, new strategies for the filtering or data assimilation and prediction of moisture coupled tropical waves are needed and this is the topic of the present paper.

The approach proposed here is to address the issues of data assimilation and prediction through a suite of analog models which despite their simplicity capture key features of the observational record and physical processes in moisture coupled tropical waves. This approach is analogous to the use of various versions of the Lorenz-96 model (Lorenz 1996; Wilks 2005; Majda *et al.* 2005; Abramov and Majda 2007; Crommelin and Vanden-Eijnden 2008; Harlim and Majda 2008a, 2010a; Majda and Harlim 2012, and references therein) to gain insight into basic issues for midlatitude filtering, prediction, and parameterization. The viability of this approach for moisture coupled tropical waves rests on

recent advances in simplified modelling of convectively coupled tropical waves and the MJO which predict key physical features of these waves such as their phase speed, dispersion relation, front to rear tilt (Kiladis *et al.* 2005, 2009), and circulation in qualitative agreement with observations (Khouider and Majda 2006a,b, 2007, 2008b,a; Majda *et al.* 2007; Majda and Stechmann 2009a,b, 2011) through simplified moisture-coupled models. The analog models emphasized here involve the multicloud convective parameterization based on three cloud types congestus, deep, and stratiform, above the boundary layer (Khouider and Majda 2006a,b, 2007, 2008b,a). The convective closure of the multicloud model takes into account the energy available for congestus and deep convection and uses a nonlinear moisture switch that allows for natural transitions between congestus and deep convection as well as for stratiform downdrafts which cool and dry the boundary layer. As a simplified two vertical baroclinic mode model, the multicloud model is very successful in capturing most of the spectrum of convectively coupled waves (Kiladis *et al.* 2009; Khouider and Majda 2008a) as well as the nonlinear organization of large scale envelopes mimicking across scale interactions of the MJO and convectively coupled waves (Khouider and Majda 2007, 2008b). Furthermore, the multicloud parameterization has been used in the next generation NCAR-GCM (HOMME) and is very successful in simulating the MJO and convectively coupled equatorial waves, at a coarse resolution of 170km in the idealized case of a uniform SST (aquaplanet) setting (Khouider *et al.* 2011). A stochastic version of the multicloud model has been utilized recently as a novel convective parameterization to improve the physical variability of deficient deterministic convective parameterizations (Khouider *et al.* 2010; Frenkel *et al.* 2011b).



The filtering skill for the recovery of troposphere moisture, heating profiles, precipitation, and vertical tilts in circulation and temperature from sparse noisy partial observations is studied here for a turbulent MJO-like travelling wave (Majda *et al.* 2007) and for the temporal development of a convectively coupled wave train. A suite of filters with judicious model errors, based on linear stochastic models (Harlim and Majda 2008a, 2010a; Majda and Harlim 2012) in a moisture coupled eigenmode basis is developed here and applied to the two test problems as well as related 3D-VAR algorithms with a full moist background covariance matrix or a dry background covariance (Zagar *et al.* 2004b,a). These results are the first demonstration of the utility of the analog multicloud models for gaining insight for data assimilation and prediction of moisture coupled tropical waves.

The plan for the remainder of the paper is the following. In Section 2, the suite of simplified tropical models for filtering and prediction is reviewed; section 3 illustrates two simplified cases, an MJO analog wave (Majda *et al.* 2007) and the temporal development of a convectively coupled tropical wave train which illustrate phenomena in the models and also serve as examples for filtering in subsequent sections of the paper. The suite of filters with judicious model errors for moisture coupled tropical waves are introduced in Section 4. Filtering skill for these algorithms applied to the MJO analog wave and the development of a convectively coupled wave train is reported in Section 5. Section 6 is a concluding discussion and summary.

## 2. Test models with moisture coupled tropical waves

The test models proposed here begin with two coupled shallow water systems: a direct heating mode forced by a bulk precipitation rate from deep penetrative clouds (Neelin and Zeng 2000) and a second vertical baroclinic mode forced by the upper level heating (cooling) and

lower level cooling (heating) of stratiform and congestus clouds, respectively (Khouider and Majda 2006a). Below, for simplicity in exposition, we present these equations without explicit nonlinear advection effects and coupling to barotropic winds. This allows us to emphasize moisture coupled tropical waves here but we comment later in this section about how nonlinear advection and barotropic winds enrich the dynamics of the test models. Thus, the test models begin with two equatorial shallow water equations

$$\begin{aligned} \frac{\partial \mathbf{v}_j}{\partial t} + \bar{\mathbf{U}} \cdot \nabla \mathbf{v}_j + \beta y \mathbf{v}_j^\perp - \theta_j &= -C_d u_0 \mathbf{v}_j - \frac{1}{\tau_w} \mathbf{v}_j, \\ \frac{\partial \theta_1}{\partial t} + \bar{\mathbf{U}} \cdot \nabla \theta_1 - \text{div } \mathbf{v}_1 &= P + S_1, \\ \frac{\partial \theta_2}{\partial t} + \bar{\mathbf{U}} \cdot \nabla \theta_2 - \frac{1}{4} \text{div } \mathbf{v}_2 &= -H_s + H_c + S_2. \end{aligned} \quad (1)$$

The equations in (1) are obtained by a Galerkin projection of the hydrostatic primitive equations with constant buoyancy frequency onto the first two baroclinic modes. More details of their derivation are found in (Neelin and Zeng 2000; Frierson *et al.* 2004; Stechmann and Majda 2009). In (1),  $\mathbf{v}_j = (u_j, v_j)_{j=1,2}$  represent the first and second baroclinic velocities assuming  $G(z) = \sqrt{2} \cos(\pi z/H_T)$  and  $G(2z) = \sqrt{2} \cos(2\pi z/H_T)$  vertical profiles, respectively, while  $\theta_j, j = 1, 2$  are the corresponding potential temperature components with the vertical profiles  $G'(z) = \sqrt{2} \sin(\pi z/H_T)$  and  $2G'(2z) = 2\sqrt{2} \sin(2\pi z/H_T)$ , respectively. Therefore, the total velocity field is approximated by

$$\begin{aligned} \mathbf{V} &\approx \bar{\mathbf{U}} + G(z) \mathbf{v}_1 + G(2z) \mathbf{v}_2, \\ w &\approx -\frac{H_T}{\pi} \left[ G'(z) \text{div } \mathbf{v}_1 + \frac{1}{2} G'(2z) \text{div } \mathbf{v}_2 \right], \end{aligned}$$

where  $\mathbf{V}$  is the horizontal velocity and  $w$  is the vertical velocity. The total potential temperature is given

approximately by

$$\Theta \approx z + G'(z)\theta_1 + 2G'(2z)\theta_2.$$

Here  $H_T \approx 16$  km is the height of the tropical troposphere with  $0 \leq z \leq H_T$  and  $\mathbf{v}_j^\perp = (-v_j, u_j)$  while  $\bar{\mathbf{U}}$  is the incompressible barotropic wind which is set to zero hereafter, for the sake of simplicity. In (1),  $P \geq 0$  models the heating from deep convection while  $H_s, H_c$  are the stratiform and congestus heating rates. Conceptually, the direct heating mode has a positive component and serves to heat the whole troposphere and is associated with a vertical shear flow. The second baroclinic mode is heated by the congestus clouds,  $H_c$ , from below and by the stratiform clouds,  $H_s$ , from above and therefore cooled by  $H_c$  from above and by  $H_s$  from below. It is associated with a jet shear flow in the middle troposphere (Khouider and Majda 2006a, 2007, 2008b,a). The terms  $S_1$  and  $S_2$  are the radiative cooling rates associated with the first and second baroclinic modes respectively.

The system of equations in (1) is augmented by an equation for the boundary layer equivalent potential temperature,  $\theta_{eb}$ , and another for the vertically integrated moisture content,  $q$ .

$$\begin{aligned} \frac{\partial \theta_{eb}}{\partial t} &= \frac{1}{h_b}(E - D), \\ \frac{\partial q}{\partial t} + \bar{\mathbf{U}} \cdot \nabla q + \text{div}(\mathbf{v}_1 q + \tilde{\alpha} \mathbf{v}_2 q) & \\ + \tilde{Q} \text{div}(\mathbf{v}_1 + \tilde{\lambda} \mathbf{v}_2) &= -\frac{2\sqrt{2}}{\pi} P + \frac{1}{H_T} D. \end{aligned} \quad (2)$$

In (2),  $h_b \approx 500$  m is the height of the moist boundary layer while  $\tilde{Q}$ ,  $\tilde{\lambda}$ , and  $\tilde{\alpha}$  are parameters associated with a prescribed moisture background and perturbation vertical profiles. According to the first equation in (2),  $\theta_{eb}$  changes in response to the downdrafts,  $D$ , and the sea surface evaporation  $E$ . Here the term downdraft

refers to the subsiding air resulting from evaporative cooling of congestus clouds and stratiform rain in the middle of the troposphere which therefore results in the moistening of the middle troposphere (increasing  $q$ ) and drying and cooling the boundary layer by bringing low  $\theta_e$  from aloft. The troposphere moisture equation for  $q$  is derived from the bulk water vapor budget equation by imposing a moisture stratification-like background vertical profile  $q_v = Q(z) + q$ . A detailed pedagogical derivation of this equation starting from the equations of bulk cloud microphysics is presented in Khouider and Majda (2006b). The approximate numerical values of  $\tilde{\lambda} = 0.8$  and  $\tilde{\alpha} = 0.1$ , follow directly from the derivation, while the coefficient  $\tilde{Q}$  arises from the background moisture gradient. We use the standard value  $\tilde{Q} \approx 0.9$  (Neelin and Zeng 2000; Frierson *et al.* 2004).

In full generality, the parametrizations in (1) and (2) automatically have conservation of an approximation to vertically integrated moist static energy. Notice that, the precipitation rate in (2), balances the vertical average of the total convective heating rate in (1), therefore leading to the conservation of the vertical average of the equivalent potential temperature  $\langle \theta_e \rangle = \langle Q(z) \rangle + q + \langle \Theta \rangle + \frac{h_b}{H_T} \theta_{eb}$  when the external forces, namely, the radiative cooling rates,  $S_1, S_2$ , and the evaporative heating,  $E$ , are set to zero. Also note that the sensible heating flux has been ignored in (1) for simplicity since this is a relatively small contribution in the tropics. Here and elsewhere in the text  $\langle f \rangle = (1/H_T) \int_0^{H_T} f(z) dz$ .

The equations in (1) and (2) for the prognostic variables  $q, \theta_{eb}, \theta_j, \mathbf{v}_j, j = 1, 2$ , are written in non-dimensional units where the equatorial Rossby deformation radius,  $L_e \approx 1,500$  km is the length scale, the first baroclinic dry gravity wave speed,  $c \approx 50$  ms<sup>-1</sup>, is the velocity scale,  $T = L_e/c \approx 8$  h is the associated time scale, and the dry-static stratification  $\bar{\alpha} = \frac{H_T N^2 \theta_0}{\pi g} \approx 15$  K is the temperature



Table I. Bulk constants in two-layer mode model.

$H_T = 16$ km	height of the tropical troposphere.
$\tilde{Q} = 0.9$	moisture stratification factor
$\tilde{\lambda} = 0.8$	baroclinic contribution to the moisture convergence associated with the moisture background
$\tilde{\alpha} = 0.1$	baroclinic contribution to the moisture (nonlinear) convergence associated with the moisture anomalies
$\tau_W = 75$ days	Rayleigh-wind friction relaxation time
$\tau_R = 50$ days	Newtonian cooling relaxation time
$C_d = 0.001$	boundary layer turbulent momentum friction
$L_e \approx 1500$ km	equatorial deformation radius, length scale
$T = L_e/c \approx 8$ h	time scale
$\bar{\alpha} \approx 15$ K	dry static stratification, temperature scale
$N = 0.01 s^{-1}$	Brunt-Vaisala buoyancy frequency
$\theta_0 = 300$ K	reference temperature
$h_b = 500$ m	boundary layer height
$\alpha_2 = 0.1$	relative contribution of $\theta_2$ to the middle troposphere $\theta_e$

unit scale. The basic bulk parameters of the model are listed in Table I for the readers convenience.

### 2.1. The convective parameterization

The surface evaporative heating,  $E$ , in (2) obeys an adjustment equation toward the boundary layer saturation equivalent potential temperature,  $\theta_{eb}^*$ ,

$$\frac{1}{h_b} E = \frac{1}{\tau_e} (\theta_{eb}^* - \theta_{eb}), \quad (3)$$

with  $\tau_e$  is the evaporative time scale. Besides the second baroclinic moisture advection in (2), the originality of the multicloud model resides in the treatment of the deep convective heating/precipitation,  $P$ , and the downdrafts,  $D$ , as well as the introduction of the congestus heating,  $H_c$ , into the  $\theta_2$  equation. The middle tropospheric equivalent potential temperature anomaly is defined approximately by

$$\theta_{em} \approx q + \frac{2\sqrt{2}}{\pi} (\theta_1 + \alpha_2 \theta_2). \quad (4)$$

Notice that the coefficient  $2\sqrt{2}/\pi$  in (4) results from the vertical average of the first baroclinic potential temperature,  $\sqrt{2}\theta_1 \sin(\pi z/H_T)$ , while the small value for  $\alpha_2$  adds a non-zero contribution from  $\theta_2$  to  $\theta_{em}$  to include its contribution from the lower middle troposphere although

its vertical average is zero. Following Khouider and Majda (2006a, 2008b,a), we use a switch parameter  $\Lambda$  which serves as a measure for the moistness and dryness of the middle troposphere. When the discrepancy between the boundary layer and the middle troposphere equivalent potential temperature is above some fixed threshold,  $\theta^+$ , the atmosphere is defined as dry. Moist parcels rising from the boundary layer will have their moisture quickly diluted by entrainment of dry air, hence losing buoyancy and stop to convect. In this case, we set  $\Lambda = 1$  which automatically inhibits deep convection in the model (see below). When this discrepancy is below some lower value,  $\theta^-$ , we have a relatively moist atmosphere and we set  $\Lambda = \Lambda^* < 1$ . The function  $\Lambda$  is then interpolated (linearly) between these two values. More precisely, we set

$$\Lambda = \begin{cases} 1 & \text{if } \theta_{eb} - \theta_{em} > \theta^+ \\ A(\theta_{eb} - \theta_{em}) + B & \text{if } \theta^- \leq \theta_{eb} - \theta_{em} \leq \theta^+ \\ \theta^* & \text{if } \theta_{eb} - \theta_{em} < \theta^- \end{cases} \quad (5)$$

The value of  $\theta^-$  represents a threshold below which the free troposphere is locally moist and “accepts” only deep convection while the value of  $\theta^+$  defines complete dryness.

Therefore, the precipitation,  $P$ , and the downdrafts,  $D$ , obey

$$P = \frac{1 - \Lambda}{1 - \Lambda^*} P_0 \quad \text{and} \quad D = \Lambda D_0, \quad (6)$$

while the stratiform and congestus heating rate,  $H_s$  and  $H_c$ , solve the relaxation-type equations

$$\frac{\partial H_s}{\partial t} = \frac{1}{\tau_s} (\alpha_s P - H_s) \quad (7)$$

and

$$\frac{\partial H_c}{\partial t} = \frac{1}{\tau_c} \left( \alpha_c \frac{\Lambda - \Lambda^*}{1 - \Lambda^*} \frac{D}{H_T} - H_c \right), \quad (8)$$

respectively. The dynamical equations in (1), (2), (7), and (8) define the multicloud model. Notice that, as anticipated above, when the middle troposphere is dry,  $\Lambda = 1$ , deep convection is completely inhibited, even if  $P_0$ , i.e., CAPE is positive, whereas congestus heating is favored. In the absence of deep convection the downdrafts are interpreted as the subsidence associated with the detrainment of shallow clouds. In this sense the shallow clouds serve to moisten and precondition the middle troposphere to sustain deep convection by lowering  $\Lambda$  in the model via both the increase of  $q$  and the decrease of  $\theta_{eb}$ . The situation is somewhat inverted during the deep convective episodes when  $\Lambda = \Lambda^*$ . Nevertheless, when this downdraft minimum fraction is reached, the downdraft will increase because of increasing stratiform heating,  $H_s$ , and the vanishing congestus heating,  $H_c$  (because of the factor  $\Lambda - \Lambda^*$  in Eq. (8)). Moreover, the dry atmosphere increases the downdrafts,  $D$ , and promotes boundary layer clouds. This also is well reflected in the model. Other variants of the equation in (8) for  $H_c$  can be utilized where changes in  $H_c$  respond to low-level CAPE (Khouider and Majda 2008b,a).

effect

The quantities  $P_0$  and  $D_0$  represent respectively the maximum allowable deep convective heating/precipitation and downdrafts, independent of the value of the switch function  $\Lambda$ . Notice that conceptually the model is not bound to any type of convective parametrization. A Betts-Miller relaxation type parametrization as well as a CAPE parametrization can be used to setup a closure for  $P_0$ . Recall that a Betts-Miller type parametrization consists of relaxing the moisture  $q$  (and/or the temperature) toward a fixed vertical profile,  $\hat{q}$ , (typically a tropical sounding or a moist adiabat) over some convective relaxation time  $\tau_{conv}$  (Betts and Miller 1986). A CAPE parametrization, on the other hand, is based on the kinetic energy available for deep convection which is directly converted into upward motion whenever deep convection is triggered. Recall also that CAPE is computed as the vertical integral of the buoyancy of the rising moist parcel which is proportional to the difference between the boundary layer and the environmental saturation equivalent potential temperatures,  $\theta_{eb} - \theta_{eb}^*$  (Emanuel 1994). Furthermore,  $\theta_{eb}$  anomalies are often approximated by some linear function of the tropospheric dry potential temperature (e.g. Majda and Shefter 2001). Here we let

$$P_0 = \frac{1}{\tau_{conv}} \left[ a_1 \theta_{eb} + a_2 (q - \hat{q}) - a_0 (\theta_1 + \gamma_2 \theta_2) \right]^+, \quad (9)$$

where  $f^+ = \max(f, 0)$  and  $\hat{q}$  is a threshold constant value measuring a significant fraction of the tropospheric saturation and  $\tau_{conv}$ ,  $a_1$ ,  $a_2$ ,  $a_0$  are parameters specified below. In particular the coefficient  $a_0$ , which is somewhat related to the inverse buoyancy relaxation time of Fuchs and Raymond (2002), is an important parameter to vary. The parameter  $\gamma_2$ , which couples  $\theta_2$  to  $P_0$  is also varied to assess the effects of the lower troposphere temperature variation on the parametrizations; a relatively warm lower troposphere will promote evaporation and



detrainment of cumulus clouds. Thus, it should result in a weakening of the deep convection.

The downdrafts are closed by

$$D_0 = \frac{m_0}{\bar{P}} [\bar{P} + \mu_2(H_s - H_c)]^+ (\theta_{cb} - \theta_{em}), \quad (10)$$

where  $m_0$  is a scaling of the downdraft mass flux and  $\bar{P}$  is a prescribed precipitation/deep convective heating at radiative convective equilibrium. Here  $\mu_2$  is a parameter allowing for stratiform and congestus mass flux anomalies (Majda and Shefter 2001; Majda *et al.* 2004). Finally the radiative cooling rates,  $S_1, S_2$  in (1) are given by a simple Newtonian cooling model

$$S_j = -Q_{R,j}^0 - \frac{1}{\tau_R} \theta_j, \quad j = 1, 2, \quad (11)$$

where  $Q_{R,j}^0, j = 1, 2$  are the radiative cooling rates at radiative convective equilibrium (RCE). This is a spatially homogeneous steady state solution where the convective heating is balanced by the radiative cooling. The basic constants in the model convective parametrization and the typical values utilized here are given in Table II. The physical features incorporated in the multi-cloud model are discussed in detail in (Khouider and Majda 2006a, 2007, 2008b,a).

## 2.2. Moisture coupled phenomena in the test models

As already noted in the introduction, the dynamic multicloud models in (1), (2), (7), (8) capture a number of observational features of equatorial convectively coupled waves and the MJO. These phenomena occur in multi-wave dynamical models with strong moisture coupling through (2), nonlinear on-off switches like (5), (9), (10) and nonlinear saturation of moisture coupled instabilities (Khouider and Majda 2006a, 2007, 2008b,a; Khouider *et al.* 2011). All of these features present

major challenges for contemporary data assimilation and prediction ~~strategy~~. Two detailed analog examples are presented in Section 3.

*strategies.*

As described in detail in Khouider and Majda (2006b) the multicloud models also include the quasi-equilibrium models (Neelin and Zeng 2000; Frierson *et al.* 2004; Pauluis *et al.* 2008) which mimic the Betts-Miller and Arakawa-Schubert parameterizations of GCM's in a limiting regime. Such models arise formally by keeping

the first baroclinic mode in (1), retaining the moisture equation in (2) with  $D = 0$ , setting  $\Lambda = 1$  in (6), and using  $P_0$  in (9) with  $a_1 = 0$  while ignoring all remaining boundary layer and cloud equations. There are many interesting exact solutions of the nonlinear dynamics with moisture switches in ~~these~~ regime, large scale precipitation fronts, which serve as interesting test problems for filtering with nonlinear switches and moisture coupled waves (Frierson *et al.* 2004; Pauluis *et al.* 2008; Stechmann and Majda 2006); the behavior of finite ensemble Kalman filters (Evensen 1994; Anderson 2001; Bishop *et al.* 2001; Hunt *et al.* 2007) and particle filters (van Leeuwen 2010; Anderson 2010) are particularly interesting in this context with moisture coupled switches and exact solutions. However, rigorous mathematical theory establishes that these quasi-equilibrium models have no instabilities or positive Lyapunov exponents (Majda and Souganidis 2010), unlike realistic tropical convection and the full multicloud models. More realism in the quasi-equilibrium tropical models can be achieved by allowing active barotropic dynamics and coupled nonlinear advection which allows for tropical-extratropical wave interactions (Lin *et al.* 2000; Majda and Biello 2003; Biello and Majda 2004). Examples with these features are developed by Khouider and Majda (2005a,b).

*this*  
*quasi equilibrium*

**Table II.** Parameters in the convective parametrization. The parameters in the middle panel will be chosen differently for the MJO-analogue case in Section 3.1 and the temporal development of a convectively coupled wave train in Section 3.2. The parameters in the lower panel are determined at the RCE state.

$\theta_{eb}^* - \bar{\theta}_{eb} = 10 \text{ K}$	Discrepancy between boundary layer $\theta_e$ at its saturated value and at the RCE state
$\theta^\pm = 10, 20 \text{ K}$	temperature threshold used to define the switch function $\Lambda$
$A, B$	linear fitting constant interpolating the switch function $\Lambda$
$\alpha_s = 0.25$	Stratiform heating adjustment coefficient
$\gamma_2 = 0.1$	Relative contribution of $\theta_2$ to convective parametrization
$Q_{R,1}^0 = 1 \text{ K day}^{-1}$	second baroclinic radiative cooling rate
$\Lambda^*$	Lower threshold of the switch function $\Lambda$
$\mu_2$	Relative contribution of stratiform and congestus mass flux anomalies to the downdrafts
$\alpha_c$	Congestus heating adjustment coefficient
$\tau_s$	Stratiform heating adjustment time
$\tau_c$	Congestus heating adjustment time
$a_0$	Inverse buoyancy time scale of convective parametrization
$a_1$	Relative contribution of $\theta_{eb}$ to convective parametrization
$a_2$	Relative contribution of $q$ to convective parametrization
$\tau_{conv}$	Deep convection adjustment time
$\bar{\theta}_{eb} - \bar{\theta}_{em}$	discrepancy between boundary and middle troposphere potential temperature at RCE value
$\tau_e \approx 8 \text{ h}$	Evaporation time scale in the boundary layer
$Q_{R,2}^0$	second baroclinic radiative cooling rate
$\hat{q}$	threshold beyond which condensation takes place in Betts-Miller scheme
$m_0$	Scaling of downdraft mass flux

### 3. Examples of moisture coupled tropical waves in the test model

In this section, we describe two concrete examples <sup>which</sup> solutions will be used as the truth for generating synthetic observations (as we will describe in Section 4). The two specific examples include an MJO-like traveling wave (Majda *et al.* 2007) and <sup>the</sup> initiation of a convectively coupled wave train that mimics the solutions of explicit simulations with <sup>a</sup> Cloud Resolving Model (Grabowski and Moncrieff 2001). Following the basic setup in Khouider and Majda (2006a, 2007), we consider the multicloud model in (1), (2), (7), (8) on a periodic equatorial ring without rotation,  $\beta = 0$ , without barotropic wind,  $\bar{\mathbf{U}} = 0$ , and with a uniform background sea surface temperature given by constant  $\theta_{eb}^*$ . With this setup, the wind velocity in (1), (2) has only the zonal wind component,  $\mathbf{v}_j = u_j$ .

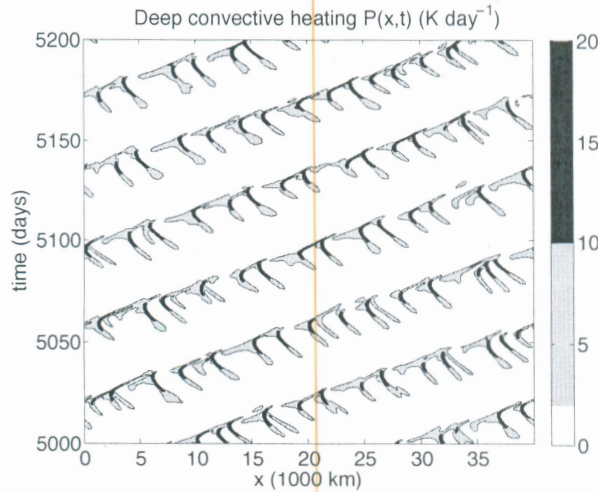
#### 3.1. An MJO-like turbulent traveling wave

In our first example, we consider the parameter regime for <sup>an</sup> intraseasonal MJO-like turbulent traveling wave. Following Majda *et al.* (2007), we set the bulk parameters in Table I,  $\bar{Q} = 1$ ,  $\bar{\lambda} = 0.6$ ,  $C_d = 10^{-5}$ ,  $\tau_W = 150$  days,  $\tau_R = 50$  days and the convective parameters in Table II,  $\bar{\theta}_{eb} - \bar{\theta}_{em} = 12 \text{ K}$ ,  $a_0 = 12$ ,  $a_1 = 0.1$ ,  $a_2 = 0.9$ ,  $\mu_2 = 0.5$ ,  $\alpha_c = 0.5$ ,  $\Lambda^* = 0.2$ . The intraseasonal timescale is generated through  $\tau_{conv} = 12$  hours which is consistent with the current observational estimates for large-scale consumption of CAPE and  $\tau_s = \tau_c = 7$  days which is also consistent with the current observational record for low-level moistening and congestus cloud development in the MJO.

The linear stability analysis for this parameter regime has been studied in detail in Majda *et al.* (2007). Here, we <sup>report</sup> some of the important features for the readers convenience: the unstable wavenumbers 2 and 3 have growth rates of roughly  $(30 \text{ days})^{-1}$  and phase speed of 6.9 and 5.8  $\text{ms}^{-1}$ , respectively. These unstable modes have westward, tilted

**SUMMARIZE**  
for eastward propagating waves





**Figure 1.** Contour plot of the deep convective heating  $P(x, t)$  from a numerical simulation of the multicloud model with parameter values in Section 3.1, Tables I, and II. Heating values of greater than  $2 \text{ K day}^{-1}$  are shaded in grey while greater than  $10 \text{ K day}^{-1}$  are shaded in black.

vertical structure for heating, velocity, and temperature, with clear first and second baroclinic mode contributions and low-level cooler potential temperature leading and within the deep convection. In Figure 1, we show the contour plot of the precipitation  $P$  (which is exactly the deep heating rate for this model) at the statistical steady state from a numerical simulation between 5000–5200 days. The main feature here is an eastward moving wavenumber-2 wave with phase speed  $6.1 \text{ ms}^{-1}$  which is in agreement with the linear analysis phase speed. Within the envelope of this wave are intense westward moving small scale fluctuations. These fluctuations occur irregularly and there are often long breaks between intense deep convective events. All of these features are observed in the MJO (Zhang 2005).

### 3.2. Initiation of a convectively coupled wave train

In this second example, we consider the three cloud model with enhanced congestus heating (Khouider and Majda 2008b) with slightly different parameterization than the above. In particular, the total precipitation,  $P$ , is different from the deep convection heating rate,  $H_d$ , and is defined as

follows,

$$P = \frac{2\sqrt{2}}{\pi} (H_d + \xi_s H_s + \xi_c H_c), \quad (12)$$

*allowing for stratiform and congestus rain.*

The key feature in this new parameterization is attributed to the asymmetric heating rate contribution in the upper and lower level atmosphere with nonzero  $\xi_s$  and  $\xi_c$ , respectively. This new feature replaces ~~only~~ the first baroclinic heating equation in (1) with

$$\frac{\partial \theta_1}{\partial t} - \frac{\partial u_1}{\partial x} = H_d + \xi_s H_s + \xi_c H_c + S_1. \quad (13)$$

The moisture equation in (2) remains unchanged except that now we remove the scale factor  $\frac{2\sqrt{2}}{\pi}$  in front of  $P$  since it is already included in (12).

The new congestus parameterization uses exactly the same switch function  $\Lambda$  in (5) with middle-troposphere equivalent potential temperature approximation in (4). The major changes replace the precipitation,  $P$ , in (6) with

$$H_d = (1 - \Lambda) Q_d, \quad (14)$$

with bulk energy available for deep convection given by

$$Q_d = \left\{ \bar{Q} + \frac{1}{\tau_{\text{conv}}} [a_1 \theta_{eb} + a_2 q - a_0 (\theta_1 + \gamma_2 \theta_2)] \right\}^+. \quad (15)$$

In (15), parameter  $\bar{Q}$  is the bulk convective heating determined at the RCE state. The downdraft in (6) is also replaced with

$$D = \frac{m_0}{Q} \left[ \bar{Q} + \mu_2 (H_s - H_c) \right]^+ (\theta_{eb} - \theta_{em}). \quad (16)$$

Compared to (6), this new parameterization assigns  $\Lambda^* = 0$  for the deep convection heating rate and ignores the factor  $\Lambda$  in the original downdraft equation. The corresponding

dynamical equations for the stratiform and congestus heating are

$$\frac{\partial H_s}{\partial t} = \frac{1}{\tau_s}(\alpha_s H_d - H_s), \quad (17)$$

$$\frac{\partial H_c}{\partial t} = \frac{1}{\tau_c}(\alpha_c \Lambda Q_c - H_c), \quad (18)$$

where

$$Q_c = \left\{ \bar{Q} + \frac{1}{\tau_{conv}} [\theta_{eb} - a'_0(\theta_1 + \gamma_2 \theta_2)] \right\}^+ \quad (19)$$

denotes a "bulk energy" for congestus heating.

In our numerical experiment, we use the same parameter values as in Khouider and Majda (2008b). The bulk constants in Table I are not changed. The convective parameters in Table II are used with  $\Lambda^* = 0$ ,  $\mu_2 = 0.25$ ,  $\alpha_c = 0.1$ ,  $\tau_s = 3$  h,  $\tau_c = 1$  h,  $a_0 = 5$ ,  $a_1 = a_2 = 0.5$ ,  $\tau_{conv} = 2$  h, and  $\bar{\theta}_{eb} - \theta_{em} = 14$  K. The additional new parameters for the enhanced congestus parameterization include the coefficients representing contributions of stratiform and congestus clouds to the first baroclinic heating,  $\xi_s = 0.5$  and  $\xi_c = 1.25$ , respectively; inverse convective buoyancy time scale associated with congestus clouds,  $a'_0 = 2$ ; the bulk convective heating  $\bar{Q}$  that is determined at RCE. Interested readers should consult Khouider and Majda (2008b) for the detail of the linear stability analysis.

Here, we are interested in the initiation of a convectively coupled wave train to mimic the two-dimensional explicit Cloud Resolving Model solutions in Grabowski and Moncrieff (2001). In particular, we integrate the model with a single wave at 20,000 km (center of the domain) as the initial condition (see the space-time plot of the first two-baroclinic velocities, potential temperatures, congestus and deep heating rates, moisture, and precipitation in Figure 2). Note that this setup is exactly

the regime analyzed in Frenkel *et al.* (2011a) in which they focused on understanding the effect of diurnal cycle and we neglect the diurnal cycle here. Notice there are fast moving waves (see  $q$ ,  $H_d$ , and  $P$  in Figure 2) during the first 2 days moving away from the 20,000 km mark. After about 8-10 days, additional waves appear; this wave initiation is partly due to the convectively coupled wave interactions with faster moving gravity waves. After about 100 days, these waves mature to a wave train of six individual eastward moving waves with a velocity of approximately  $14.5 \text{ ms}^{-1}$  (see Figure 3). Such wave structure and wave train organization resemble the structure found in the explicit simulations with cloud resolving model by Grabowski and Moncrieff (2001). Moreover, the matured waves have a total convective heating pattern (with backward and upward tilt in the wind and temperature fields, upper-tropospheric warm temperature anomalies slightly leading the region of the upward motion, which is in phase with the heating anomalies, with low level convergence) that is very similar to convectively coupled Kelvin waves observed in nature (Wheeler and Kiladis 1999; Wheeler *et al.* 2000; Straub and Kiladis 2002).

#### 4. Algorithm for filtering moisture coupled waves from sparse observations

In this section, we first describe the sparse observation networks and then discuss in details the reduced stochastic filtering algorithms.

##### 4.1. Sparse observation networks

In the present paper, we consider horizontally sparse observations at every 2,000 km. This means we only have  $M = 20$  observations at  $x_j = jh$ ,  $h = 2\pi/40,000$  km in a non-dimensionalized unit assuming that the equatorial belt circumference is 40,000 km. For compact notation, we define  $\Psi_{j,m} = (u_1, u_2, \theta_1, \theta_2, \theta_{eb}, q, H_s, H_c)^T$ ; we use

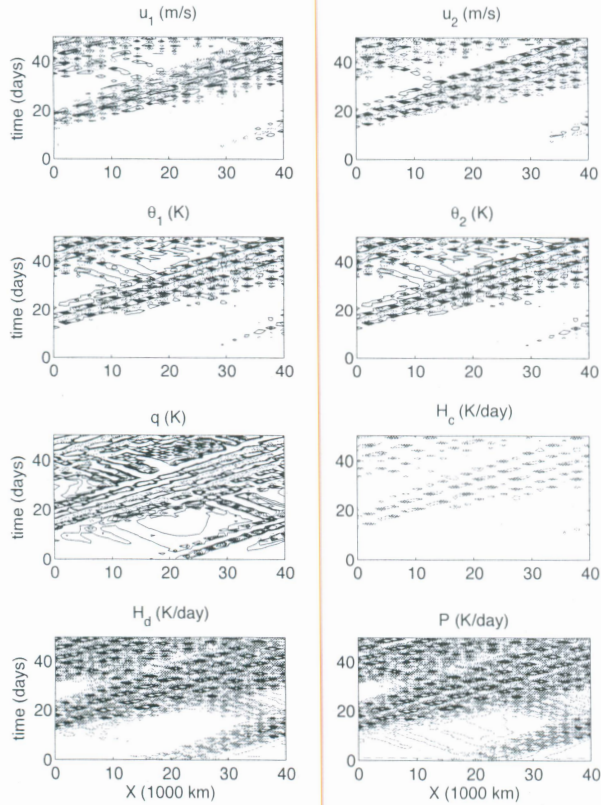
high resolution

localized piece of a single unstable linear wave of small amplitude centered at 20,000 km.

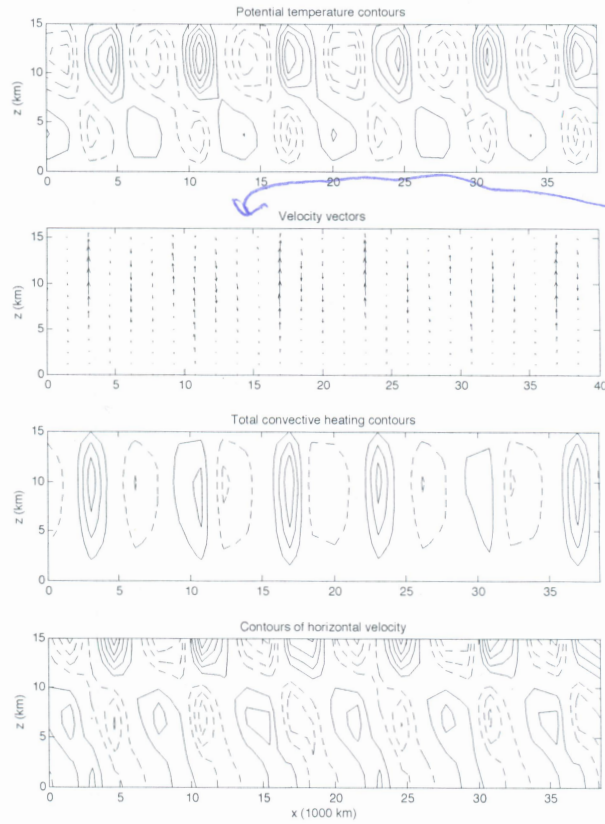
wave speed

with





**Figure 2.** Initiation of a convectively coupled wave train: the space-time plot here is constructed with coarse spatial and temporal resolutions at every 2000 km and 24h (This coarse data set is sampled from solutions with higher resolutions at every 40 km and 3 hours). The contour intervals are  $0.25 \text{ ms}^{-1}$  for the zonal wind,  $0.025 \text{ K}$  for the potential temperature and humidity, and  $0.05 \text{ K day}^{-1}$  for the heating rates and precipitation. Solid black (dash grey) contours denote positive (negative) values for  $u_1, u_2, \theta_1, \theta_2, q$ . Solid black (dash grey) contours denote heating rates greater (smaller) than  $1 \text{ K day}^{-1}$  for  $H_c, H_d, P$ .



**Figure 3.** Moving average of the vertical structure in a reference frame of  $14.5 \text{ ms}^{-1}$  from time period of 500–1000 days. The contour intervals are  $0.07 \text{ K}$  for the potential temperature,  $0.54 \text{ K day}^{-1}$  for the total convective heating, and  $0.35 \text{ ms}^{-1}$  for the horizontal velocity. Solid (dashes) contours denote positive (negative) values.

subscripts  $j$  and  $m$  to specify that each component in  $\Psi$  is evaluated at grid point  $x_j$  and discrete time  $t_m$ , respectively. We define a general observation model

$$\mathbf{G}\Psi_{j,m}^o = \mathbf{G}\Psi_{j,m} + \mathbf{G}\sigma_{j,m}, \quad \sigma_{j,m} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}^o), \quad (20)$$

where  $\mathbf{G}$  is an observation operator that maps the model state to the observation state space and  $\sigma_{j,m}$  are eight-dimensional independent Gaussian white noises with mean zero and diagonal covariance matrix  $\mathbf{R}^o$ . Vertically, we consider four observation networks with specific  $\mathbf{G}$ :

**SO (Surface Observations):** Here, we consider observing the wind, potential temperature at a surface height  $z_s = 100 \text{ m}$ , and the equivalent boundary layer potential temperature  $\theta_{eb}$ . The corresponding observation operator is a  $3 \times 8$  matrix  $\mathbf{G}$  with nonzero components

$$\begin{aligned} \mathbf{G}_{1,1} &= G(z_s), \mathbf{G}_{1,2} = G(2z_s), \\ \mathbf{G}_{2,3} &= G'(z_s), \mathbf{G}_{2,4} = 2G'(2z_s), \\ \mathbf{G}_{3,5} &= 1, \end{aligned} \quad (21)$$

where  $G, G'$  are the vertical baroclinic profiles defined in Section 2.

**SO+MT (Surface Observations + Middle Troposphere Temperature):** This observation network includes temperature at middle-troposphere height  $z_m = 8$  km in addition to SO. The corresponding observation operator is a  $4 \times 8$  matrix  $\mathbf{G}$  with nonzero components

$$\mathbf{G}_{4,3} = G'(z_m), \mathbf{G}_{4,4} = 2G'(2z_m), \quad (22)$$

in addition to (21).

**SO+MTV (Surface Observations + Middle Troposphere Temperature & Velocity):** This observation network includes velocity at middle-troposphere height  $z_m = 8$  km in addition to SO+MT. The corresponding observation operator is a  $5 \times 8$  matrix  $\mathbf{G}$  with nonzero components

$$\mathbf{G}_{5,1} = G(z_m), \mathbf{G}_{5,2} = 2G(2z_m). \quad (23)$$

in addition to (21) and (22).

**CO (Complete Observations):** This vertically complete observation network is defined with  $\mathbf{G} = \mathbf{I}$  for diagnostic purpose. *(purposes)*

#### 4.2. Filtering Algorithms

In this paper, we consider the simplest version of our reduced stochastic filters, the Mean Stochastic Model (MSM, Harlim and Majda 2008a, 2010a,b; Majda and Harlim 2012). The new feature in the present context is that we have multiple variables  $\Psi_j$  as opposed to a scalar field and therefore we need to design the MSM in an appropriate coordinate expansion to avoid parameterizing various coupling terms.

As in Harlim and Majda (2008a), our design of the filter prior model is attributed to the standard approach for modeling turbulent fluctuations (Majda *et al.* 1999; Majda and Timofeyev 2004; DelSole 2004; Majda *et al.*

2008), that is: We introduce model errors through linearizing the nonlinear models about a frozen constant state and replacing the truncated nonlinearity with a dissipation and spatially correlated noise (white in time) to mimic rapid energy transfer between different scales. In the present context, we consider the linearized multicloud model about the RCE,

$$\frac{d\Psi'}{dt} = \mathcal{P}(\partial_x)\Psi', \quad (24)$$

where  $\Psi'$  denotes the perturbation field about the RCE and  $\mathcal{P}$  denotes the linearized differential operator of the multicloud model at RCE. A comprehensive study of the linear stability analysis of (24) involves solving eigenvalues of an  $8 \times 8$  dispersion matrix,  $\omega(k)$ , *and* was reported in Majda *et al.* (2007) for the MJO-like wave and in Khouider and Majda (2008b); Frenkel *et al.* (2011a) for the multicloud model with enhanced congestus heating.

Consider a numerical discretization for (24) with spatial mesh size of  $\Delta x = 2000$  km such that the model state space is essentially similar to the observation state space. With this approximation, the PDE in (24) becomes

$$\frac{d\hat{\Psi}_k}{dt} = i\omega(k)\hat{\Psi}_k, \quad |k| \leq M/2 = 10, \quad (25)$$

where  $\{\hat{\Psi}_k\}_{|k| \leq M/2}$  are the discrete Fourier components of  $\{\Psi'_j\}_{j=1,\dots,M}$ . Now consider an eigenvalue decomposition,  $i\omega(k)\mathbf{Z}_k = \mathbf{Z}_k\mathbf{\Lambda}_k$ , where  $\mathbf{\Lambda}_k$  is a diagonal matrix of the eigenvalues and  $\mathbf{Z}_k$  is a matrix whose columns are the corresponding eigenvectors. Then we can write (25) as a diagonal system,

$$\frac{d\hat{\Phi}_k}{dt} = \mathbf{\Lambda}_k\hat{\Phi}_k, \quad |k| \leq M/2 = 10, \quad (26)$$

*based on*

with the following transformation

$$\hat{\Phi}_k = \mathbf{Z}_k^{-1} \hat{\Psi}_k. \quad (27)$$

#### 4.2.1. The MSM-Filter

The Mean Stochastic Model is defined through the following stochastic differential system,

$$d\hat{\Phi}_k = [(-\Gamma_k + i\Omega_k)\hat{\Phi}_k + \mathbf{f}_k]dt + \Sigma_k dW_k, \quad (28)$$

for  $|k| \leq M/2$ . In (28),  $\Gamma_k$ ,  $\Omega_k$ , and  $\Sigma_k$  are diagonal matrices with diagonal components obtained through regression fitting to the climatological statistics while the forcing term is proportional to the climatological mean field,  $\mathbf{f}_k = (\Gamma_k - i\Omega_k)\langle\hat{\Phi}_k\rangle$ ; here, the angle bracket  $\langle\cdot\rangle$  denotes an average. Notice that the realizability of this stochastic model (referred as MSM-I Majda *et al.* 2010; Harlim and Majda 2010b; Majda and Harlim 2012) is guaranteed since  $\Gamma_k$  is always positive definite as opposed to the alternative approach which sets  $\Omega_k = -i\Lambda_k$  (Penland 1989; DelSole 2000).

The discrete-time Kalman filtering problem with the MSM as the prior model is defined for each horizontal wavenumber  $k$  as follows

$$\hat{\Psi}_{k,m} = \mathcal{F}_k(\Delta t)\hat{\Psi}_{k,m-1} + \mathbf{g}_{k,m} + \eta_{k,m}, \quad (29)$$

$$\mathbf{G}\hat{\Psi}_{k,m}^o = \mathbf{G}\hat{\Psi}_{k,m} + \mathbf{G}\hat{\sigma}_{k,m}, \quad (30)$$

where the observation model in (30) is the discrete Fourier component of the canonical observation model in (20) with Gaussian noises,  $\hat{\sigma}_{k,m} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}^o/M)$ . The discrete filter

model in (29) has coefficients

$$\mathcal{F}_k(t) = \mathbf{Z}_k \exp\left((-\Gamma_k + i\Omega_k)t\right)\mathbf{Z}_k^{-1}, \quad (31)$$

$$\mathbf{g}_{k,m} = -(\mathbf{I} - \mathcal{F}_k(t_m))(-\Gamma_k + i\Omega_k)^{-1}\mathbf{f}_k, \quad (32)$$

and unbiased Gaussian noises  $\eta_{k,m}$  with covariance matrix

$$\mathbf{Q}_k = \frac{1}{2}\mathbf{Z}_k \Sigma_k^2 \Gamma_k^{-1} (\mathbf{I} - |\mathcal{F}_k(\Delta t)|^2) \mathbf{Z}_k^*. \quad (33)$$

These coefficients are obtained by evaluating the analytical solutions of the stochastic differential system in (28) at observation time interval  $\Delta t = t_{m+1} - t_m$  and applying the transformation in (27).

The MSM-filter in (29)-(30) is computationally very cheap since it only involves  $M/2 + 1$  independent  $8 \times 8$  Kalman filtering problems, ignoring cross-correlations between different horizontal wavenumbers. Such a diagonal approximation may seem to be counterintuitive since it generates severe model errors but we have shown that it provides high filtering skill beyond the perfect model simulations in various context<sup>s</sup> including the regularly spaced sparse observations (Harlim and Majda 2008b), irregularly spaced sparse observations (Harlim 2011), strongly chaotic nonlinear dynamical systems (Harlim and Majda 2008a, 2010a), and midlatitude ~~can~~ <sup>baroclinic</sup> wave dynamics (Harlim and Majda 2010b).

Applying the Kalman filter formula on each wavenumber in (29)-(30) provides the following background (or prior) mean and error covariance estimates,

$$\hat{\Psi}_{k,m}^b = \mathcal{F}_k(\Delta t)\hat{\Psi}_{k,m-1}^a + \mathbf{g}_{k,m} \quad (34)$$

$$\mathbf{R}_{k,m}^b = \mathcal{F}_k(\Delta t)\mathbf{R}_{k,m-1}^a \mathcal{F}_k(\Delta t)^* + \mathbf{Q}_k, \quad (35)$$



and analysis (or posterior) mean and error covariance estimates

$$\begin{aligned}\hat{\Psi}_{k,m}^a &= \hat{\Psi}_{k,m}^b + \mathbf{K}_{k,m}(\mathbf{G}\hat{\Psi}_{k,m}^o - \mathbf{G}\hat{\Psi}_{k,m}^b) \\ \mathbf{R}_{k,m}^a &= (\mathbf{I} - \mathbf{K}_{k,m}\mathbf{G})\mathbf{R}_{k,m}^b, \\ \mathbf{K}_{k,m} &= \mathbf{R}_{k,m}^b\mathbf{G}^*(\mathbf{G}(\mathbf{R}_{k,m}^b + \mathbf{R}^o/M)\mathbf{G}^*)^{-1},\end{aligned}\quad (36)$$

where  $\mathbf{K}_{k,m}$  is the Kalman gain matrix.

#### 4.2.2. The complete 3D-VAR

For diagnostic purpose, we also consider a 3D-VAR version in the MSM framework above. That is, we simply set the background error covariance matrix to be independent of time,

$$\mathbf{B}_k \equiv \lim_{\Delta t \rightarrow \infty} \mathbf{R}_{k,m}^b = \frac{1}{2}\mathbf{Z}_k\mathbf{\Sigma}_k^2\mathbf{\Gamma}_k^{-1}\mathbf{Z}_k^*, \quad (37)$$

and repeat the mean prior and posterior updates in (34), (36) with a constant Kalman gain matrix,

$$\mathbf{K}_k = \mathbf{B}_k\mathbf{G}^*(\mathbf{G}(\mathbf{B}_k + \mathbf{R}^o/M)\mathbf{G}^*)^{-1}.$$

We called this approach the complete 3D-VAR because the forward model parameters in (31), (32), (33) and the background covariance matrix in (37) are determined from complete solutions of the multicloud model in (1), including the ~~moist~~ <sup>moisture</sup> and heating variables from (2), (7), (8). This formulation is significantly different from earlier approach with variational technique (Zagar *et al.* 2004b,a) in which the background covariance matrix is parameterized in an eigenmode basis constructed from dry equatorial waveguide.

#### 4.2.3. The “dry and cold” 3D-VAR

To mimic the approach in (Zagar *et al.* 2004b,a), we consider only using the wind and temperature data,  $u_1, u_2, \theta_1, \theta_2, \theta_{eb}$ , to construct the “dry and cold” eigenmode basis and background covariance matrix  $\mathbf{B}_k$ . Technically, we still use the MSM model in (28) but replace the transformation in (27) with

$$\hat{\Phi}_k^{dc} = \mathbf{Z}_k^{-1} \begin{bmatrix} \mathbf{I}_{5 \times 5} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \hat{\Psi}_k. \quad (38)$$

In this sense, the parameters  $\mathbf{\Gamma}_k$ ,  $\mathbf{\Omega}_k$ , and  $\mathbf{\Sigma}_k$  in (28) are fitted to climatological statistics of  $\hat{\Phi}_k^{dc}$  based on only the wind and temperature variables. Repeating the 3D-VAR algorithm described above in this setup provides an honest “dry and cold” version analogous to the earlier approach in Zagar *et al.* (2004b,a).

Besides the eigenmode basis difference, we should note that the “dry and cold” 3D-VAR here is computationally much cheaper than that in Zagar *et al.* (2004b,a) since we perform both the prior and posterior updates in the diagonalized Fourier basis with reduced stochastic filters through (34)-(36) as opposed to their approach that propagates the nonlinear dry shallow water equations in the physical space and applies the analysis step in the spectral diagonal basis. On each data assimilation step, their approach requires back-and-forth transformations in between the physical and spectral spaces with a rotational transformation matrix that is quite often ill-conditioned as reported in Zagar *et al.* (2004b).

## 5. Filtering skill for moisture coupled tropical waves

In this section, we report the numerical results of implementing the filtering algorithms in Section 4.2 to assimilate the synthetic sparse observation networks defined in Section 4.1 on the two examples discussed in Section 3.

In the numerical simulation <sup>§</sup> below, we consider the precise observations case with  $\mathbf{R}^o = \mathbf{0}$  and small observation noises with positive definite covariance matrix  $\mathbf{R}^o > \mathbf{0}$ . In the latter case, we choose the noise variances to be less than both the peak of the energy spectrum and the smallest average signal amplitude. For ex, the peak of the energy spectrum in variable  $u_1$  is roughly  $10^{-3}$  and the smallest average signal amplitude (in Fourier domain) is about  $10^{-5}$ , and we choose the first diagonal component of the noise covariance matrix to be  $\mathbf{R}_{11}^o/M = 2 \times 10^{-6}$  (here,  $M$  is the normalization factor from the discrete Fourier transform).

### 5.1. MJO-like turbulent traveling wave

Our goal here is to check the filtering skill in recovering the climatological structure of the MJO-like traveling wave (Section 3.1) with the MSM-forward model in (28) with parameters, (31)-(33), which are specified from a timeseries of 2000 days at the climatological state.

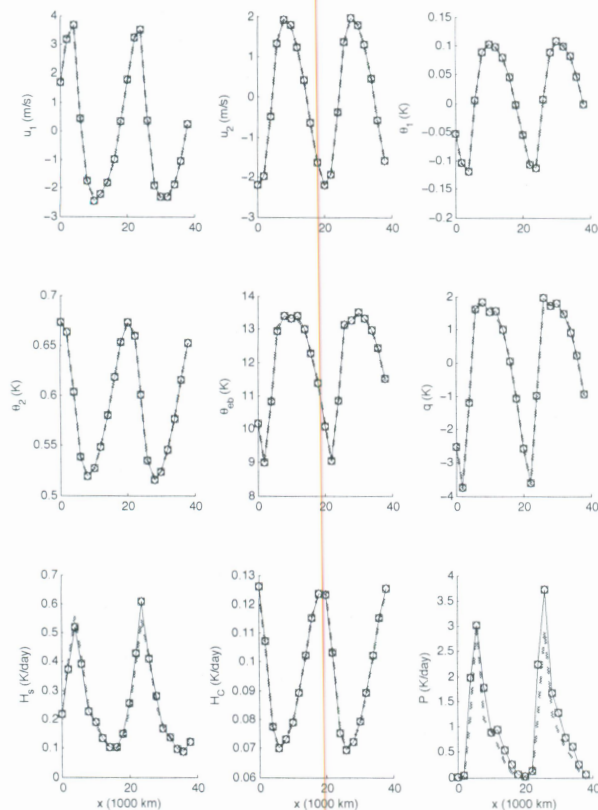
First, we compare the moving average of  $u_1, u_2, \theta_1, \theta_2, \theta_{eb}, q, H_s, H_c, P$  obtained from the true solutions of the test model in Section 3.1 and the posterior mean estimates in (36). The moving average is taken in a reference frame at  $6.1 \text{ ms}^{-1}$  from time period of 750-1000 days and we refer to the resulting moving average profile as the "climatological structure". In Figures 4-8, we show the moving average from assimilations with observation time interval of 24 hours for complete observations (CO) with  $\mathbf{R}^o = \mathbf{0}$ , and for all observation networks discussed in Section 4.1, CO, SO+MTV, SO+MT, SO with small observation noise covariance,  $\mathbf{R}^o > \mathbf{0}$ . For observation network CO without observation errors,  $\mathbf{R}^o = \mathbf{0}$  (see Figure 4), the three schemes, MSM-filter, Complete 3D-VAR and "cold and dry" 3D-VAR, are identical and they perfectly recover the climatological structure except for slight overestimation on the stratiform heating and precipitation.

In the presence of observation noise, we find that all the three schemes are able to recover  $u_1$  and  $\theta_{eb}$  with any observation network. When middle-troposphere wind observation is absent (see SO+MT and SO in Figures 7, 8), the estimate for  $u_1$  slightly degrades but for  $u_2$  is completely wrong. The MSM-filter overestimates  $\theta_2$  roughly by 0.1 K even with surface and middle-troposphere potential temperature observations; this poor estimation is attributed to inaccurate mean estimate (on the zeroth horizontal mode) of  $\theta_2$ . Both the MSM-filter and the Complete 3D-VAR are able to recover the climatological structure of the moisture  $q$  with any observation network (with slight errors for the MSM-filter with SO). On the other hand, the "cold and dry" 3D-VAR cannot produce  $q$  accurately even with observation network CO and simply predicts dry atmosphere (with zero moisture profile) when the moisture is unobserved. All the three filters are not able to reproduce the stratiform and congestus heating profiles when they are not observed.

Except for the surface observation (SO) network, both the MSM-filter and Complete 3D-VAR are able to reasonably recover the precipitation rate ( $P$ ) which in this model is exactly the deep convection heating rate; here, the "cold and dry" 3D-VAR precipitation estimate is very inaccurate (see Figures 5-7). For the surface observation (SO) network (see Figure 8), the overestimation on the precipitation can be explained as follows. From the precipitation budget in (9), it is obvious that the contributions of  $\theta_{eb}, q$ , and  $\theta_2$  to the convective parameterization are small (with scale factors  $a_1 = 0.1, a_2 = 0.5, a_0\gamma_2 = 1.2$ , respectively) relative to  $\theta_1$  (with scale factor  $a_0 = 12$ ). Therefore, the wet climate (with large precipitation estimates as seen in Figure 8) is attributed to the slight underestimation of the first baroclinic potential temperature,  $\theta_1$ . The Complete 3D-VAR underestimates  $\theta_1$  by as much as 0.5 K; this yields spatially uniform precipitation rate of about  $2.3 \text{ K day}^{-1}$ . The MSM-filter underestimates  $\theta_1$  by as much as 1.5 K

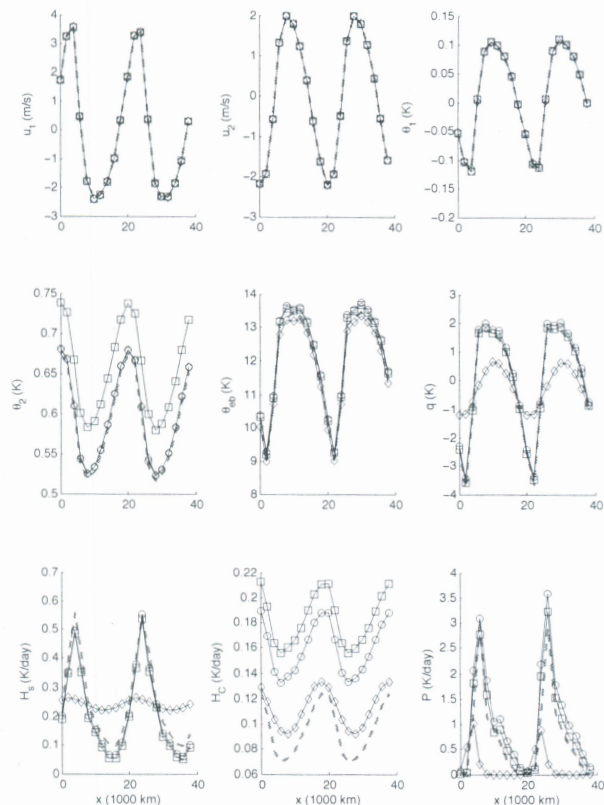


and its corresponding precipitation estimate is about 20 K day<sup>-1</sup>.



**Figure 4.** MJO-like waves with  $\Delta t = 24$  hours,  $R^o = 0$  and complete observations (CO): Moving average is in a reference frame at  $6.1 \text{ ms}^{-1}$  of the model variables. True (grey dashes), posterior mean state of the complete 3D-VAR (circles), MSM-filter (squares), and the "dry and cold" 3D-VAR (diamonds).

In Figures 9-12, we show the detail vertical structure of the total potential temperature  $\Theta$ , the velocity vector field  $(V, w)$ , the total convective heating, and horizontal velocity from the MJO-like wave in Section 3.1 and the Complete 3D-VAR estimates with observation networks SO+MTV, SO+MTV, and SO, respectively. In particular, the vertical tilted structure in the potential temperature is recovered with any of these three observation networks; similar recovery (not shown) is also obtained with the MSM-filter; the "cold and dry" 3D-VAR also recovers this tilted structure except with observation network SO. On the other hand, the tilted structure in the horizontal velocity with low level convergence that is in phase with



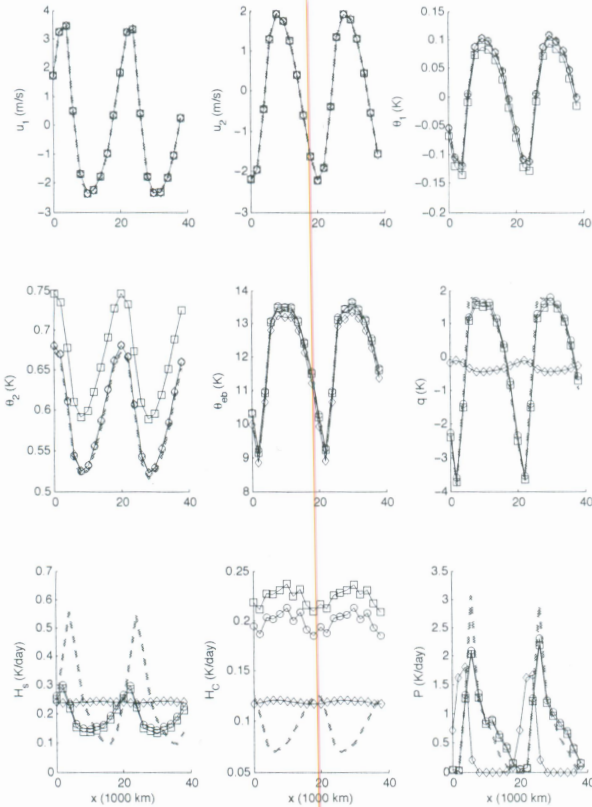
**Figure 5.** MJO-like waves with  $\Delta t = 24$  hours,  $R^o > 0$  and complete observations (CO): Moving average is in a reference frame at  $6.1 \text{ ms}^{-1}$  of the model variables. True (grey dashes), posterior mean state of the complete 3D-VAR (circles), MSM-filter (squares), and the "dry and cold" 3D-VAR (diamonds).

the deep convective heating is not recovered whenever the middle-troposphere wind observation is absent. Notice also that the deep convective heating is recovered except with observation network SO; similar recovery (not shown) is also attained with the MSM-filter but not with the "cold and dry" 3D-VAR.

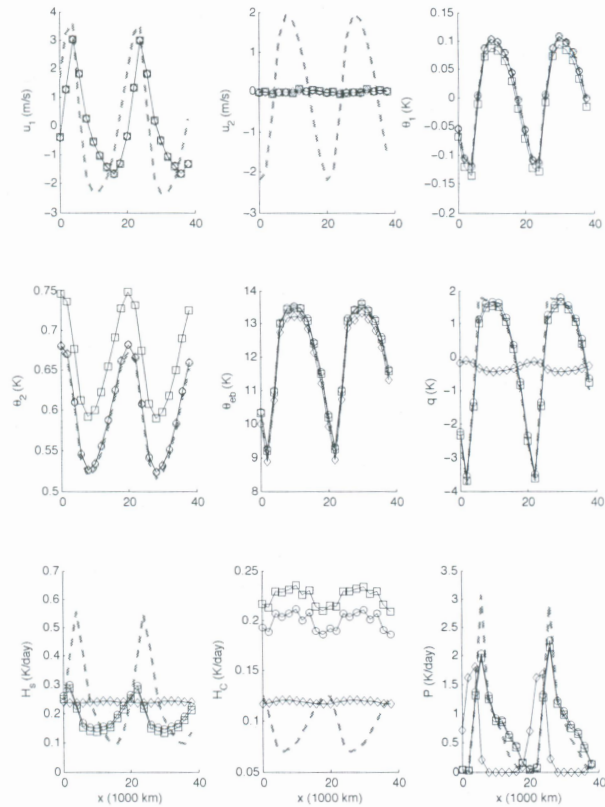
We also find that both the Complete 3D-VAR and MSM-filter are able to reconstruct the detail precipitation structure in Figure 1 except when assimilated with observation network SO (results are not shown). The "cold and dry" 3D-VAR reproduces the eastward MJO-like signal but fails to capture the westward intermittent moist fluctuations within the MJO envelope as shown in Figure 1.

We also numerically repeat the numerical experiments above with different observation time intervals ranging from





**Figure 6.** MJO-like waves with  $\Delta t = 24$  hours,  $R^o > 0$  and surface observations plus middle-troposphere potential temperature and velocity (SO+MTV): Moving average is in a reference frame at  $6.1 \text{ ms}^{-1}$  of the model variables. True (grey dashes), posterior mean state of the complete 3D-VAR (circles), MSM-filter (squares), and the “dry and cold” 3D-VAR (diamonds).



**Figure 7.** MJO-like waves with  $\Delta t = 24$  hours,  $R^o > 0$  and surface observations plus middle-troposphere potential temperature (SO+MT): Moving average is in a reference frame at  $6.1 \text{ ms}^{-1}$  of the model variables. True (grey dashes), posterior mean state of the complete 3D-VAR (circles), MSM-filter (squares), and the “dry and cold” 3D-VAR (diamonds).

6 to 96 hours with the Complete 3D-VAR and MSM-filter. Particularly noteworthy is that the posterior estimates have roughly similar RMS errors for the observed variables independent of the observation times; for the unobserved variables, the RMS errors for the shorter observation times are larger than those for the longer observation times. This latter result can be understood as follows. The dynamical operator  $\mathcal{F}_k$  in (31) is essentially ~~practically~~ marginally stable (with largest eigenvalue 0.9899) for  $\Delta t = 6$  hours and is strictly stable (with largest eigenvalue 0.8836) for longer  $\Delta t = 72$  hours. The observability condition, which is a necessary condition for accurate filtered solutions when the dynamical operator is marginally stable (Majda and Harlim 2012), is practically violated here; our test with SO+MT observation network suggests that the observability matrix

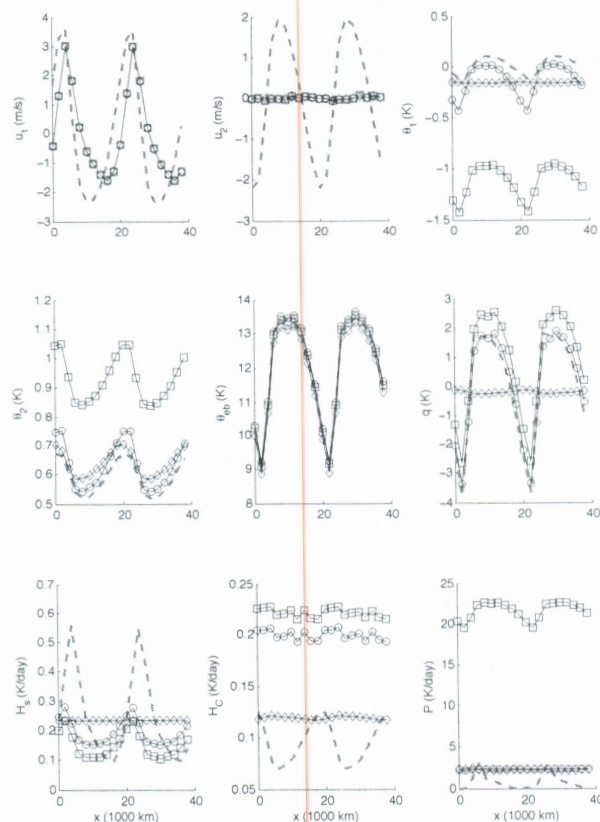
is ill-conditioned with  $\det \left( [\mathbf{G}^T (\mathbf{G} \mathcal{F}_k)^T] \right) \approx 10^{-20}$ . This explains why the longer observation times produce more accurate filtered solutions.

*We encounter similar behavior of filtered solutions in the next example.*

### 5.2. Initiation of a convectively coupled wave train

Here, our goal is to check the filtering skill in recovering the transient behavior of initiation of a convectively coupled wave train (Section 3.2) with the MSM forward model in (28) ~~with~~ <sup>where</sup> parameters, (31)–(33), are specified from a timeseries at the climatological state ~~from~~ <sup>for the</sup> period of time 500–1000 days.

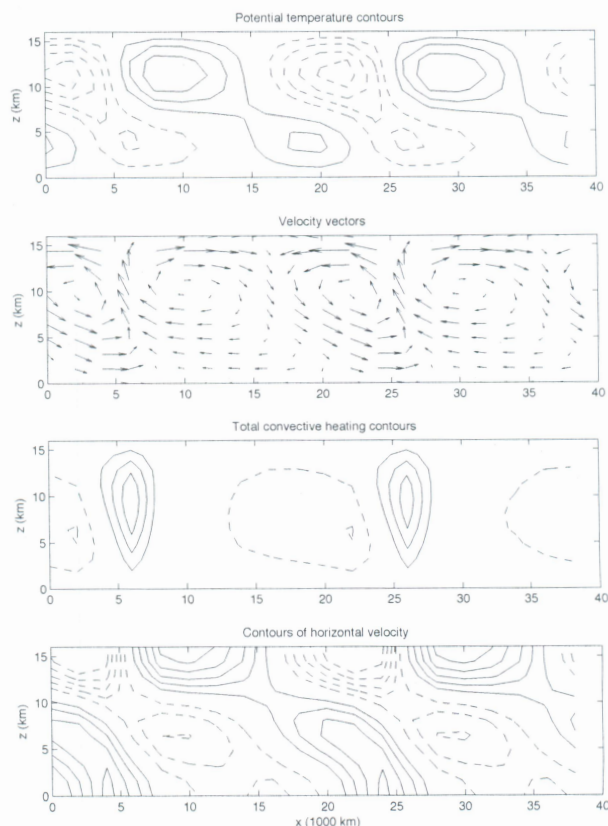
In Figures 13–16, we report the space-time plot of the filtered estimates at period of time 0–50 days from the Complete 3D-VAR with observation time  $\Delta t = 24$  hours, observation noise variance  $R^o > 0$ , and observation



**Figure 8.** MJO-like waves with  $\Delta t = 24$  hours,  $R^o > 0$  and surface observations (SO): Moving average is in a reference frame at  $6.1 \text{ ms}^{-1}$  of the model variables. True (grey dashes), posterior mean state of the complete 3D-VAR (circles), MSM-filter (squares), and the “dry and cold” 3D-VAR (diamonds).

networks SOMTV, SOMT, and SO. By eye-sight, we can see that the emerging pattern in Figure 2 is recovered for all variables except for the deep convection heating rate with complete observation network! This poor estimate is attributed to an overestimation of  $\theta_1$  (which sets the available convective heating  $Q_d$  in (15) to zero). In this case, the precipitation is completely contributed by the stratiform and congestus heating rates. On the other hand, even if the pattern of  $H_d$  is always captured with networks SO+MTV, SO+MT, SO, its accuracy is questionable as we will see below.

To be more precise, we quantify the filter skill with the average RMS error and pattern correlation (between the posterior mean estimate and the truth) at the initiation period of time 0-75 days before these waves lock into a



**Figure 9.** The true vertical profile of the MJO-like waves computed with moving average is in a reference frame at  $6.1 \text{ ms}^{-1}$ . The contour intervals are  $0.07 \text{ K}$  for the potential temperature,  $0.29 \text{ K day}^{-1}$  for the total convective heating, and  $1 \text{ ms}^{-1}$  for the horizontal velocity. Solid (dashes) contours denote positive (negative) values.

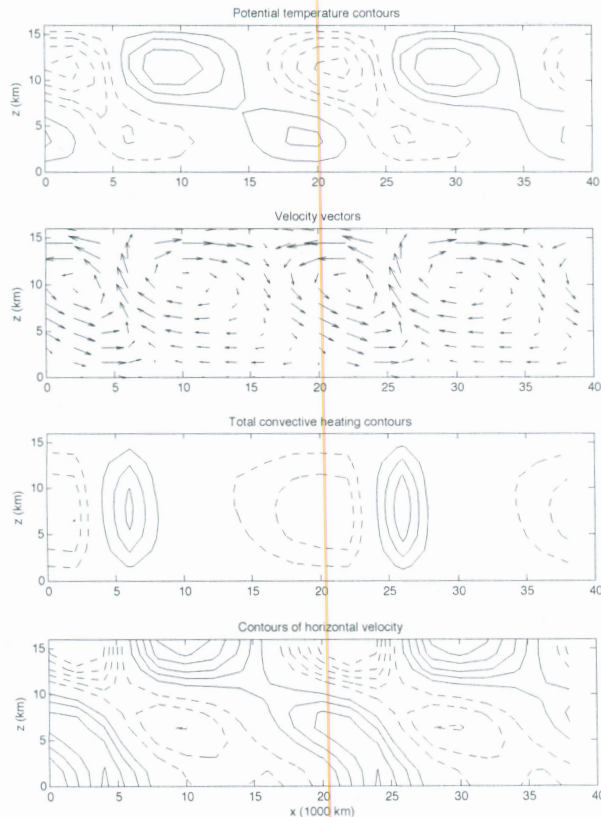
wave train of six waves as shown in Figure 3. In Figures 17-24, we plot these two performance measures as functions of observation times for observation networks CO, SO+MTV, SO+MT, and SO, respectively. In each panel, we compare four numerical experiments including the MSM-filter with  $R^o = 0$  (grey dashes) and  $R^o > 0$  (grey dashes with circle markers), the Complete 3D-VAR with  $R^o = 0$  (black solid line) and  $R^o > 0$  (black dashes with square markers).

From the average RMS errors (see Figures 17, 19, 21, 23), we find that the filtering skill of the MSM-filter and the Complete 3D-VAR are <sup>not</sup> different at all except for the wind variables when the middle-troposphere wind is not observed and  $R^o > 0$ ; there, the RMS errors of the Complete 3D-VAR are smaller than those of the MSM-Filter (see Figures 21, 23) but their pattern correlations are identical

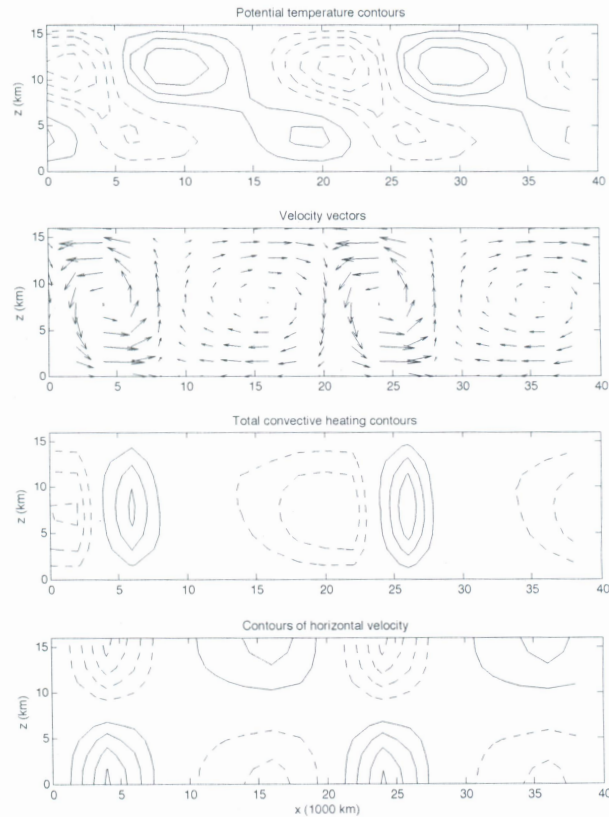
the budget in the filtered solution is dominated by the stratiform and congestus heating rates.

and not





**Figure 10.** The vertical profile from Complete 3D-VAR estimate with observation network SO+MTV and  $R^o > 0$ , and  $\Delta t = 24$  hours. The contour details are similar to those in Fig 9.



**Figure 11.** The vertical profile from Complete 3D-VAR estimate with observation network SO+MT and  $R^o > 0$ , and  $\Delta t = 24$  hours. The contour details are similar to those in Fig 9.

(see Figures 22, 24). When observations are complete (CO) and  $R^o = 0$ , both the MSM-filter and Complete 3D-VAR are able to recover the truth (with negligible errors and pattern correlation one).

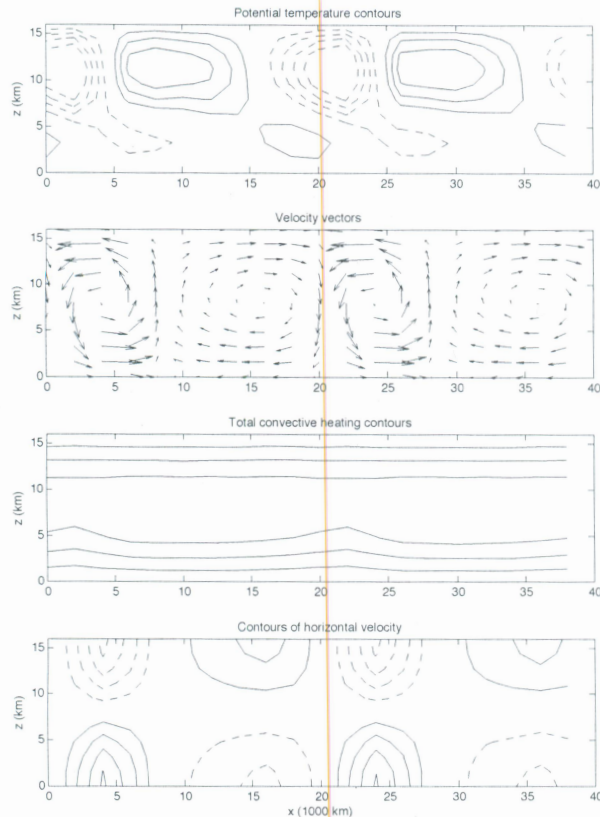
The average RMS errors for simulations with  $R^o > 0$  (dashes with markers) for variables  $\theta_1, \theta_2, q, H_c$  decay as functions of observation time even with complete observation network (see Figure 17). The larger errors with shorter observation times here are attributed to the violation of practical controllability (Majda and Harlim 2012) which is also a necessary condition for optimal filtering when the system is marginally stable (here  $\mathcal{F}_k$  has maximum eigenvalue close to 1). Additionally, we observe a similar error decaying pattern as function of observation time with SO+MTV, SO+MT, and SO for the unobserved variables  $q, H_c, H_d$ , and  $P$  even when the observed wind and potential

*We find that*

temperatures have no errors ( $R^o = 0$ ). Here, the larger errors in the unobserved variables for shorter observation times are attributed to the violation of practical observability as explained in Section 5.1.

When  $R^o > 0$ , the RMS errors of the deep convection heating rate  $H_d$  are roughly  $1 \text{ K day}^{-1}$  (see Figure 17) with observation network CO but the pattern correlations (PC) are roughly zero (the PC curves are below 0.5 in Figure 18). The PC confirms the inability to recover  $H_d$  as shown in Figure 13. Both filtering schemes with the other observation networks (SO+MTV, SO+MT, SO) recover the structure of  $H_d$  (with PC of roughly 0.6 from Figures 20, 22, 24) but their errors are very large (as much as 10 K from Figures 19, 21, 23). The failure to even capture the deep convection pattern with CO is attributed to overestimation

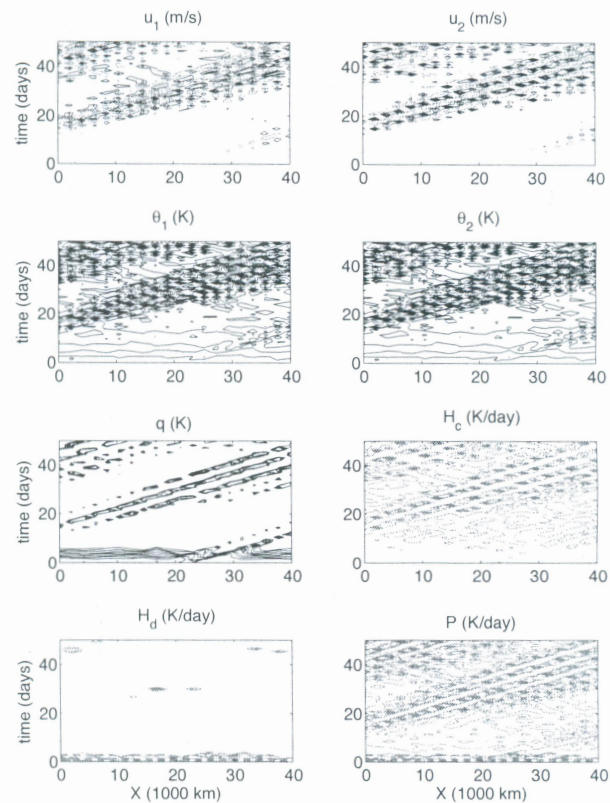




**Figure 12.** The vertical profile from Complete 3D-VAR estimate with observation network SO and  $\mathbf{R}^o > 0$ , and  $\Delta t = 24$  hours. The contour details are similar to those in Fig 9.

of  $\theta_1$  as explained before <sup>by</sup> contrasting the detail space-time structure of  $\theta_1$  in Figures 2 and 13. The average RMS errors show such <sup>a</sup> failure tendency with larger error with CO compared to those with the other networks, but they don't inform us whether the potential temperature estimates are warmer or colder than the truth which is important for accurate precipitation estimation. <sup>for failure</sup>

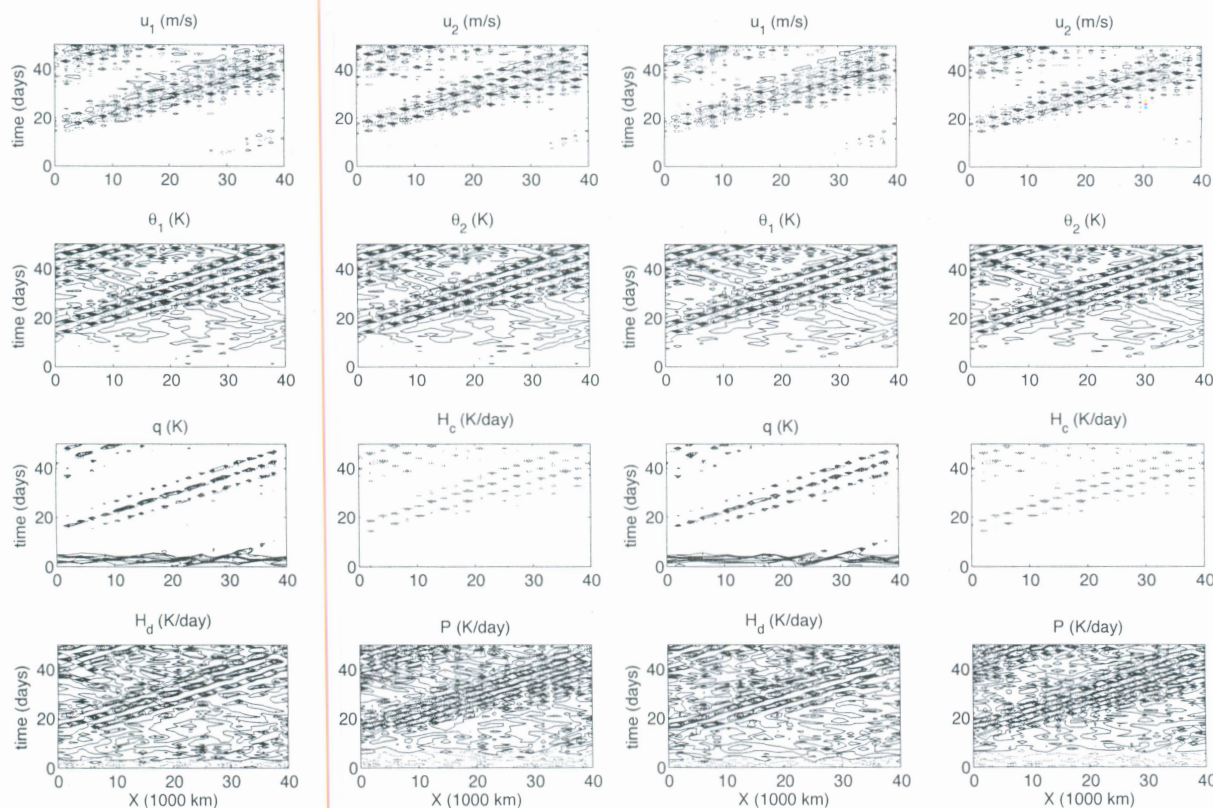
Finally, notice that with networks SO+MTV, SO+MT, and SO, the RMS errors of the unobserved variables for assimilation with  $\mathbf{R}^o = 0$  are larger than those with  $\mathbf{R}^o > 0$ ; for example, see the errors for variables  $\theta_1, q, H_c, H_d, P$  in Figure 23. <sup>can</sup> These larger errors with  $\mathbf{R}^o = 0$  are due to ill-conditioned Kalman gain matrix in (36) with sparse observation networks with operator,  $\mathbf{G} \in \mathbb{R}^{S \times 8}, S < 8$ . <sup>We find that</sup>



**Figure 13.** Space-time plot from the Complete 3D-VAR estimate with observation network CO and  $\mathbf{R}^o > 0$ , and  $\Delta t = 24$  hours. The contour intervals are  $0.25 \text{ ms}^{-1}$  for the zonal wind, the temperature,  $0.025 \text{ K}$  for the potential temperature and humidity, and  $0.05 \text{ K day}^{-1}$  for the heating rates and precipitation. Solid black (dash grey) contours denote positive (negative) values for  $u_1, u_2, \theta_1, \theta_2, q$ . Solid black (dash grey) contours denote heating rates greater (smaller) than  $1 \text{ K day}^{-1}$  for  $H_c, H_d, P$ .

## 6. Summary and concluding discussion

In this paper, we use multicloud models (Khouider and Majda 2006a, 2007; Majda *et al.* 2007; Khouider and Majda 2008b) as the test models for filtering moist tropical convection. In particular, we aim to establish guidelines for future design of filtering schemes in assimilating and predicting tropical atmospheric dynamics. We view the multicloud model, <sup>the</sup> which convective parameterizations <sup>with</sup> includes three cloud types, congestus, deep, and stratiform, above the boundary layer, as a candidate for the simplest toy model for <sup>which</sup> the tropical waves (analogous to the Lorenz 96 for the midlatitude weather dynamics) for the following reasons: It is very successful in capturing most of the spectrum of the convectively coupled <sup>moisture-coupled</sup>



**Figure 14.** Space-time plot from the Complete 3D-VAR estimate with observation network SO+MTV and  $R^o > 0$ , and  $\Delta t = 24$  hours. The contour intervals are  $0.25 \text{ ms}^{-1}$  for the zonal wind, the temperature,  $0.025 \text{ K}$  for the potential temperature and humidity, and  $0.05 \text{ K day}^{-1}$  for the congestus heating rate, and  $0.25 \text{ K day}^{-1}$  for the deep convective heating and precipitation. Solid black (dash grey) contours denote positive (negative) values for  $u_1, u_2, \theta_1, \theta_2, q$ . Solid black (dash grey) contours denote heating rates greater (smaller) than  $1 \text{ K day}^{-1}$  for  $H_c, H_d, P$ .

waves (Kiladis *et al.* 2009; Khouider and Majda 2008a) as well as the nonlinear organization of large scale envelopes mimicking across scale interactions of the MJO (our first example in Section 3.1) and convectively coupled waves (our second example in Section 3.2). More importantly, this model also captures the vertical profile with front and rear tilting, the phase speed, and dispersion relations that match the observational record (Kiladis *et al.* 2005, 2009).

Here, we demonstrate the filtering skill with a suite of reduced stochastic filters with model errors, based on linear stochastic models (Harlim and Majda 2008a, 2010a; Majda and Harlim 2012) in capturing the intraseasonal MJO-like wave (Majda *et al.* 2007) and the transient initiation of a convectively coupled wave train that

**Figure 15.** Space-time plot from the Complete 3D-VAR estimate with observation network SO+MT and  $R^o > 0$ , and  $\Delta t = 24$  hours. The contour details are similar to those in Fig 14.

resembles the results from ~~simulating the more elaborate~~ Cloud Resolving Model (Grabowski and Moncrieff 2001).

From these numerical experiments, we find the following facts: (1) The key factor for accurate precipitation estimate is an accurate estimation of the first baroclinic potential temperature. Our test problems suggests that slight overestimation in  $\theta_1$  produces dry atmosphere with no rain at all and slight underestimation in  $\theta_1$  produces wet atmosphere with ~~high~~ high precipitation rate; (2) Our simple reduced stochastic filters are able to recover moisture and precipitation field profile (even when online observations of these variables are not available) provided that the filter forward prior model is designed in a moisture coupled eigenmode basis. This result suggests that the future design of tropical data assimilation algorithm should account moisture coupled



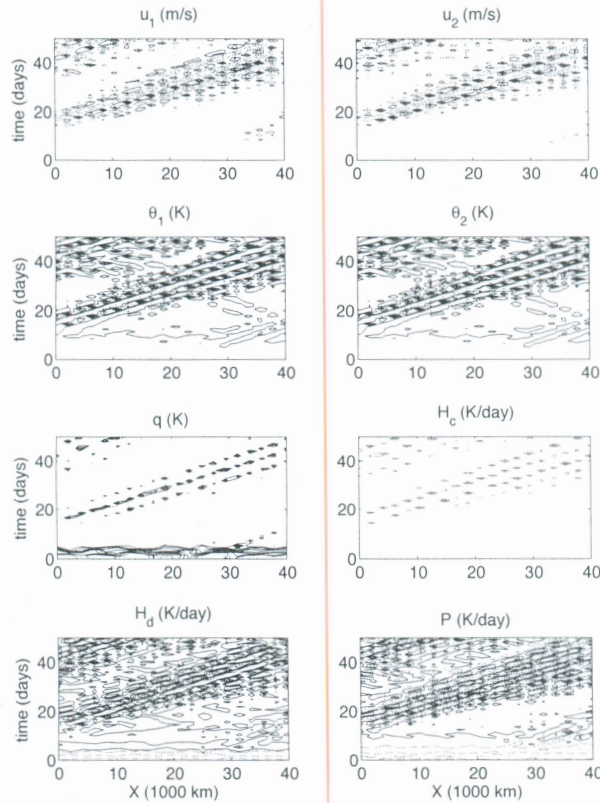


Figure 16. Space-time plot from the Complete 3D-VAR estimate with observation network SO and  $R^o > 0$ , and  $\Delta t = 24$  hours. The contour details are similar to those in Fig 14.

eigenmode basis instead of dry eigenmode basis as in Zagar *et al.* (2004b,a); (3) A better estimate for the tropical convection wave patterns requires more than surface wind and potential temperature observations; (4) The skill of the reduced filtering methods with horizontally and vertically sparse observations suggests that more accurate filtered solutions are achieved with less frequent observation times. Such a counterintuitive finding is justified with the classical observability and controllability conditions which are necessary for optimal filtering especially when the observation timescale is too short relative to the timescale of the true signal.

We hope that the encouraging results in this paper can convince researchers that are interested in the tropical data assimilation to investigate: (1) The potential of improving the estimate accuracy with more sophisticated

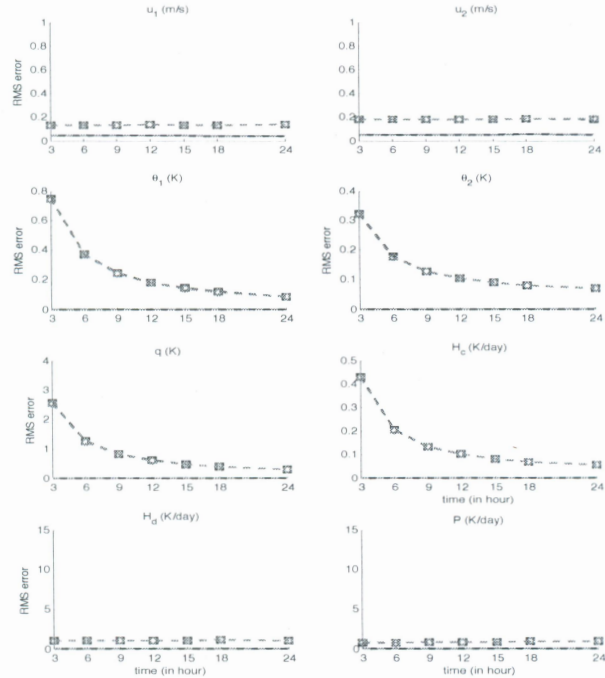


Figure 17. RMS errors as functions of observation time interval for observation network CO. MSM-filter (grey), Complete 3D-VAR (black),  $R^o = 0$  (solid line and dashes without any markers),  $R^o > 0$  (dashes with square/circle markers).

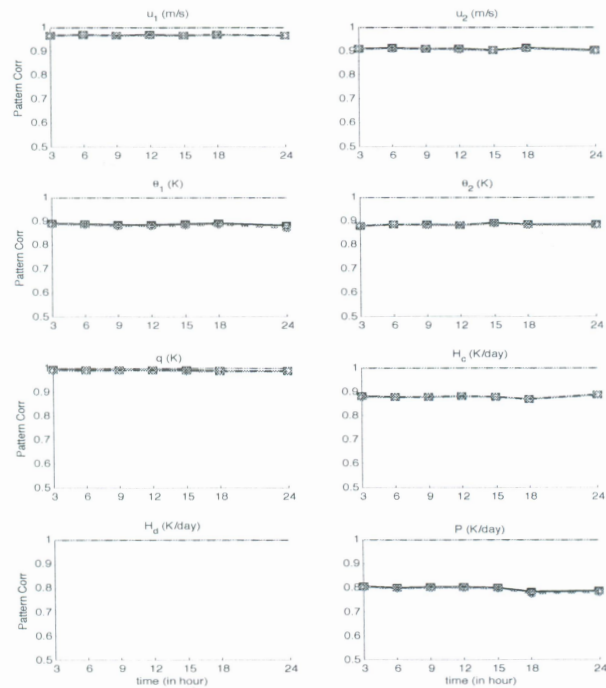
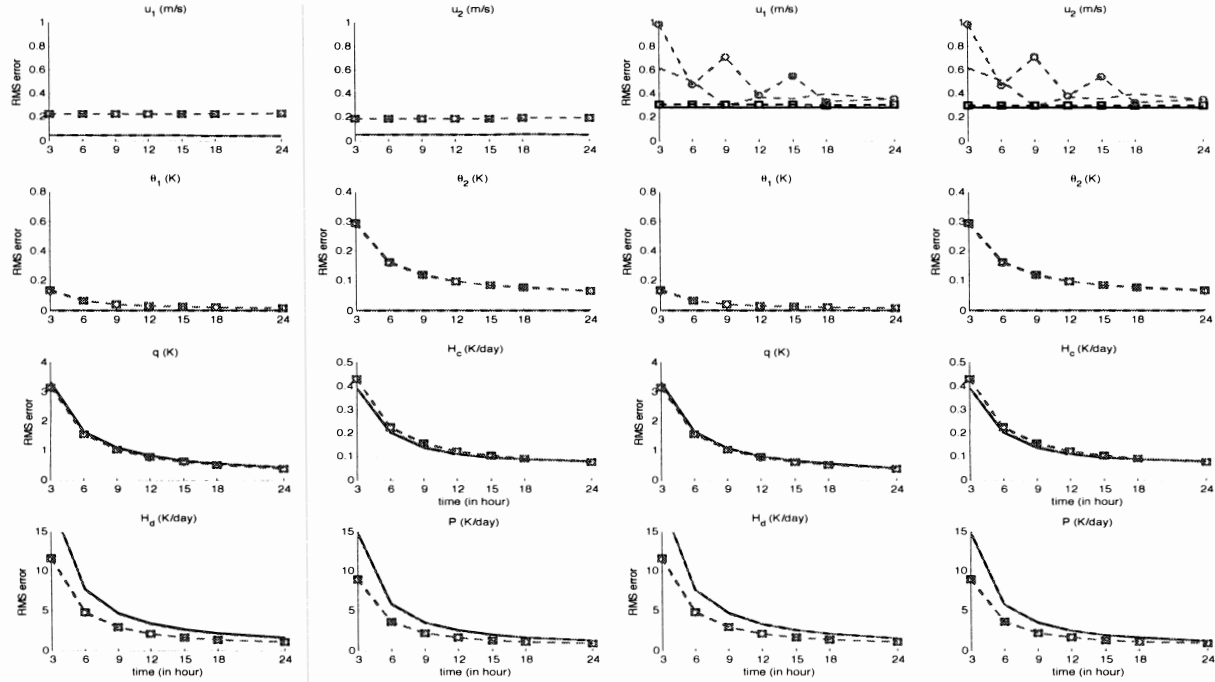


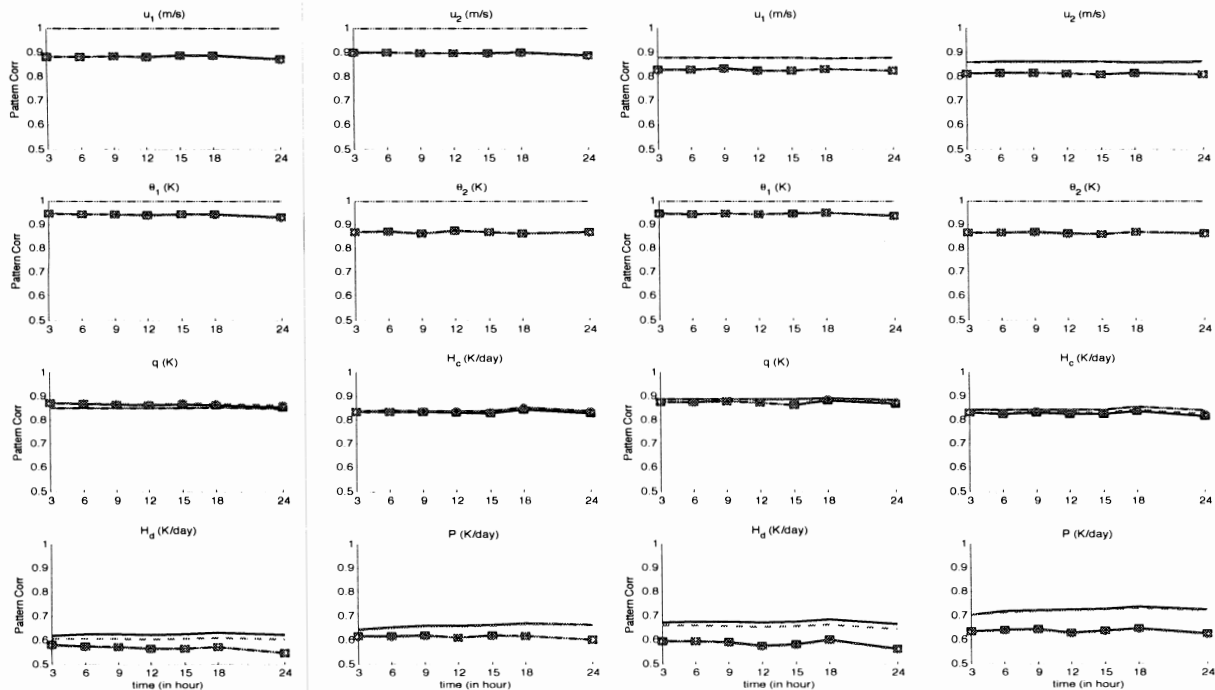
Figure 18. Pattern correlations as functions of observation time interval for observation network CO. MSM-filter (grey), Complete 3D-VAR (black),  $R^o = 0$  (solid line and dashes without any markers),  $R^o > 0$  (dashes with square/circle markers).





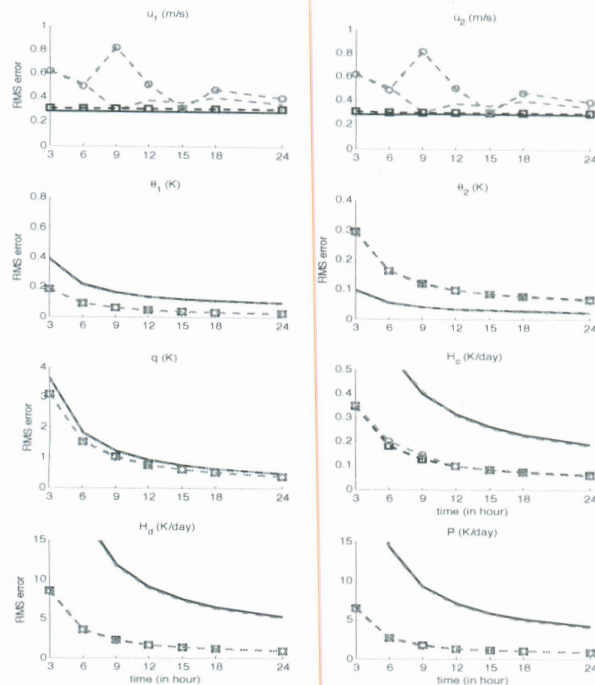
**Figure 19.** RMS errors as functions of observation time interval for observation network SO+MTV. MSM-filter (grey), Complete 3D-VAR (black),  $R^o = 0$  (solid line and dashes without any markers),  $R^o > 0$  (dashes with square/circle markers).

**Figure 21.** RMS errors as functions of observation time interval for observation network SO+MT. MSM-filter (grey), Complete 3D-VAR (black),  $R^o = 0$  (solid line and dashes without any markers),  $R^o > 0$  (dashes with square/circle markers).

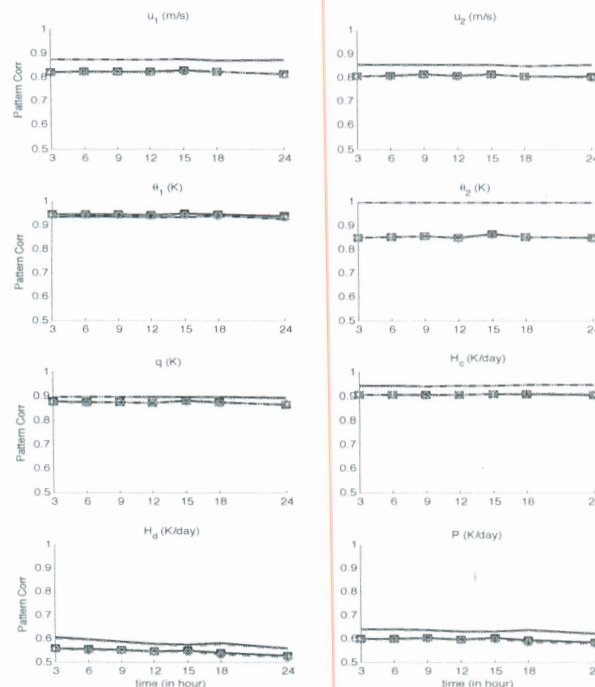


**Figure 20.** Pattern correlations as functions of observation time interval for observation network SO+MTV. MSM-filter (grey), Complete 3D-VAR (black),  $R^o = 0$  (solid line and dashes without any markers),  $R^o > 0$  (dashes with square/circle markers).

**Figure 22.** Pattern correlations as functions of observation time interval for observation network SO+MT. MSM-filter (grey), Complete 3D-VAR (black),  $R^o = 0$  (solid line and dashes without any markers),  $R^o > 0$  (dashes with square/circle markers).



**Figure 23.** RMS errors as functions of observation time interval for observation network SO. MSM-filter (grey), Complete 3D-VAR (black),  $R^o = 0$  (solid line and dashes without any markers),  $R^o > 0$  (dashes with square/circle markers).



**Figure 24.** Pattern correlations as functions of observation time interval for observation network SO. MSM-filter (grey), Complete 3D-VAR (black),  $R^o = 0$  (solid line and dashes without any markers),  $R^o > 0$  (dashes with square/circle markers).

filtering schemes through the test models here; this includes testing with the ensemble Kalman filters (Evensen 1994; Anderson 2001; Bishop *et al.* 2001; Hunt *et al.* 2007), particle filters with small ensemble sizes (van Leeuwen 2010; Anderson 2010), and other reduced stochastic filters (Gershgorin *et al.* 2010b,a; Majda and Harlim 2012) or the Gaussian closure filter (Branicki *et al.* 2011) with stochastic parameterizations that account model errors “on-the-fly”; (2) Exploring the filtering skill for other parameter regimes in the multicloud models with more realistic sea surface temperature profile, active barotropic dynamics and coupled nonlinear advection which allows for tropical-extratropical wave interactions (Lin *et al.* 2000; Majda and Biello 2003; Biello and Majda 2004); (3) Using the test models to design appropriate filters to cope with various observation networks with irregularly spaced sparse observations (Harlim 2011) and satellite measurements.

### Acknowledgement

The authors thank Boualem Khouider, Samuel Stechmann, and Yevgeniy Frenkel for sharing the relevant multicloud model source codes. The research of J.H. is partially supported by the Office of Naval Research Grant N00014-11-1-0310, the NC State startup fund, and the NC State Faculty Research and Professional Development fund. The research of A.J.M. is partially supported by the National Science Foundation Grant DMS-0456713 and the Office of Naval Research Grants ONR DRI N00014-10-1-0554 and N00014-11-7-0306.

### References

- Abramov R, Majda A. 2007. Blended response algorithm for linear fluctuation-dissipation for complex nonlinear dynamical systems. *Nonlinearity* **20**: 2793–2821.
- Anderson J. 2001. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review* **129**: 2884–2903.

- Anderson JL. 2010. A non-gaussian ensemble filter update for data assimilation. *Monthly Weather Review* **138**(11): 4186–4198, doi: 10.1175/2010MWR3253.1.
- Anderson and Moore book
- Betts AK, Miller MJ. 1986. A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, ATEX and arctic airmass data sets. *Q. J. Roy. Met. Soc.* **112**: 693–709.
- Biello JA, Majda AJ. 2004. The effect of meridional and vertical shear on the interaction of equatorial baroclinic and barotropic rossby waves. *Studies in Applied Mathematics* **112**(4): 341–390, doi:10.1111/j.0022-2526.2004.01518.x.
- Bishop C, Etherton B, Majumdar S. 2001. Adaptive sampling with the ensemble transform Kalman filter part I: the theoretical aspects. *Monthly Weather Review* **129**: 420–436.
- Branicki M, Gershgorin B, Majda A. 2011. Filtering skill for turbulent signals for a suite of nonlinear and linear extended Kalman filters. *J. Comput. Phys.* (in press) doi:doi:10.1016/j.jcp.2011.10.029.
- Crommelin D, Vanden-Eijnden E. 2008. Subgrid-scale parameterization with conditional Markov chains. *J. Atmos. Sci.* **65**: 2664–2675.
- DelSole T. 2000. A fundamental limitation of markov models. *Journal of the Atmospheric Sciences* **57**(13): 2158–2168.
- DelSole T. 2004. Stochastic model of quasigeostrophic turbulence. *Surveys in Geophysics* **25**(2): 107–149.
- Emanuel KA. 1994. *Atmospheric convection*. Oxford University Press.
- Evensen G. 1994. Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* **99**: 10 143–10 162.
- Frenkel Y, Majda AJ, Khouider B. 2011a. Simple models for the diurnal cycle and convectively coupled waves. *submitted to Theoretical and Computational Fluid Dynamics*.
- Frenkel Y, Majda AJ, Khouider B. 2011b. Using the stochastic multicloud model to improve tropical convective parameterization: A paradigm example. *Journal of the Atmospheric Sciences*.
- Frierson D, Majda A, Pauluis O. 2004. Large scale dynamics of precipitation fronts in the tropical atmosphere: A novel relaxation limit. *Comm. Math. Sci.* **2**(4): 591–626.
- Fuchs E, Raymond DJ. 2002. Large-scale modes of a nonrotating atmosphere with water vapor and cloud-radiation feedbacks. *J. Atmos. Sci.* **59**: 1669–1679.
- Gershgorin B, Harlim J, Majda A. 2010a. Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation. *J. Comput. Phys.* **229**(1): 32–57.
- Gershgorin B, Harlim J, Majda A. 2010b. Test models for improving filtering with model errors through stochastic parameter estimation. *J. Comput. Phys.* **229**(1): 1–31.
- Grabowski WW, Moncrieff MW. 2001. Large-scale organization of tropical convection in two-dimensional explicit numerical simulations. *Q. J. Roy. Met. Soc.* **127**: 445–468.
- Harlim J. 2011. Interpolating irregularly spaced observations for filtering turbulent complex systems. *SIAM J. Sci. Comp.* **33**(5): 2620–2640.
- Harlim J, Majda A. 2008a. Filtering nonlinear dynamical systems with linear stochastic models. *Nonlinearity* **21**(6): 1281–1306.
- Harlim J, Majda A. 2008b. Mathematical strategies for filtering complex systems: Regularly spaced sparse observations. *Journal of Computational Physics* **227**(10): 5304–5341.
- Harlim J, Majda A. 2010a. Catastrophic filter divergence in filtering nonlinear dissipative systems. *Comm. Math. Sci.* **8**(1): 27–43.
- Harlim J, Majda A. 2010b. Filtering turbulent sparsely observed geophysical flows. *Monthly Weather Review* **138**(4): 1050–1083.
- Houze Jr RA. 2004. Mesoscale convective systems. *Rev. Geophys.* **42**: G4003+, doi:10.1029/2004RG000150.
- Hunt B, Kostelich E, Szunyogh I. 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D* **230**: 112–126.
- Khouider B, Biello JA, Majda AJ. 2010. A stochastic multicloud model for tropical convection. *Comm. Math. Sci.* **8**: 187–216.
- Khouider B, Majda A. 2006a. A simple multicloud parametrization for convectively coupled tropical waves. Part I: Linear Analysis. *Journal of the Atmospheric Sciences* **63**: 1308–1323.

doi:10.1175/JAS-D-11-0148.1



- Khouider B, Majda A. 2007. A simple multicloud parametrization for convectively coupled tropical waves. Part II: Nonlinear Simulations. *Journal of the Atmospheric Sciences* **64**: 381–400.
- Khouider B, Majda A. 2008a. Multicloud models for organized tropical convection: Enhanced congestus heating. *Journal of the Atmospheric Sciences* **65**: 895–914.
- Khouider B, Majda AJ. 2005a. A non-oscillatory balanced scheme for an idealized tropical climate model: Part I: Algorithm and validation. *Theor. Comp. Fluid Dyn.* **19**(5): 331–354.
- Khouider B, Majda AJ. 2005b. A non-oscillatory balanced scheme for an idealized tropical climate model: Part II: Nonlinear coupling and moisture effects. *Theor. Comp. Fluid Dyn.* **19**(5): 355–375.
- Khouider B, Majda AJ. 2006b. Multicloud convective parameterizations with crude vertical structure. *Theor. Comp. Fluid Dyn.* **20**: 351–375.
- Khouider B, Majda AJ. 2008b. Equatorial convectively coupled waves in a simple multicloud model. *J. Atmos. Sci.* **65**: 3376–3397.
- Khouider B, St-Cyr A, Majda A, Tribbia J. 2011. The MJO and convectively coupled waves in a coarse-resolution GCM with a simple multicloud parameterization. *Journal of the Atmospheric Sciences* **68**: 240–264.
- Kiladis G, Straub K, Haertel P. 2005. Zonal and vertical structure of the Madden-Julian oscillation. *J. Atmos. Sci.* **62**: 2790–2809.
- Kiladis GN, Wheeler MC, Haertel PT, Straub KH, Roundy PE. 2009. Convectively coupled equatorial waves. *Rev. Geophys.* **47**: RG2003. doi:10.1029/2008RG000266.
- Lau WKM, Waliser DE (eds). 2005. *Intraseasonal variability in the atmosphere–ocean climate system*. Springer, Berlin.
- Lin JL, Kiladis GN, Mapes BE, Weickmann KM, Sperber KR, Lin W, Wheeler M, Schubert SD, Del Genio A, Donner LJ, Emori S, Gueremy JF, Hourdin F, Rasch PJ, Roeckner E, Scinocca JF. 2006. Tropical intraseasonal variability in 14 IPCC AR4 climate models Part I: Convective signals. *J. Climate* **19**: 2665–2690.
- Lin JWB, Neelin JD, Zeng N. 2000. Maintenance of tropical intraseasonal variability: Impact of evaporation–wind feedback and midlatitude storms. *Journal of the Atmospheric Sciences* **57**(17): 2793–2823.
- Lorenz E. 1996. Predictability – a problem partly solved. In: *Proceedings on predictability, held at ECMWF on 4–8 September 1995*. pp. 1–18.
- Majda A, Abramov R, Grote M. 2005. *Information theory and stochastics for multiscale nonlinear systems*. CRM Monograph Series v.25, American Mathematical Society: Providence, Rhode Island, USA.
- Majda A, Franzke C, Khouider B. 2008. An Applied Mathematics Perspective on Stochastic Modelling for Climate. *Philos Transact A Math Phys Eng Sci.* **366**(1875): 2429–2455.
- Majda A, Gershgorin B, Yuan Y. 2010. Low frequency response and fluctuation-dissipation theorems: Theory and practice. *J. Atmos. Sci.* **67**: 1181–1201.
- Majda A, Harlim J. 2012. *Filtering Complex Turbulent Systems*. Cambridge University Press (in press).
- Majda A, Stechmann S, Khouider B. 2007. Madden-Julian Oscillation analog and intraseasonal variability in a multicloud model above the equator. *Proceedings of the National Academy of Sciences* **104**(24): 9919–9924.
- Majda A, Timofeyev I. 2004. Low dimensional chaotic dynamics versus intrinsic stochastic chaos: A paradigm model. *Physica D* **199**: 339–368.
- Majda A, Timofeyev I, Vanden-Eijnden E. 1999. Models for stochastic climate prediction. *Proc. Nat. Acad. Sci.* **96**: 15 687–15 691.
- Majda AJ, Biello JA. 2003. The nonlinear interaction of barotropic and equatorial baroclinic Rossby waves. *J. Atmos. Sci.* **60**: 1809–1821.
- Majda AJ, Khouider B, Kiladis GN, Straub KH, Shefter MG. 2004. A model for convectively coupled tropical waves: nonlinearity, rotation, and comparison with observations. *J. Atmos. Sci.* **61**: 2188–2205.
- Majda AJ, Shefter MG. 2001. Waves and instabilities for model tropical convective parameterizations. *J. Atmos. Sci.* **58**: 896–914.
- Majda AJ, Souganidis PE. 2010. Existence and uniqueness of weak solutions for precipitation fronts: A novel hyperbolic free boundary problem in several space variables. *Communications on Pure and Applied Mathematics* **63**(10): 1351–1361.

- Majda AJ, Stechmann SN. 2009a. A simple dynamical model with features of convective momentum transport. *J. Atmos. Sci.* **66**: 373–392.
- Majda AJ, Stechmann SN. 2009b. The skeleton of tropical intraseasonal oscillations. *Proc. Natl. Acad. Sci.* **106**(21): 8417.
- Majda AJ, Stechmann SN. 2011. Nonlinear dynamics and regional variations in the MJO skeleton. *J. Atmos. Sci.*: accepted.
- Moncrieff M, Shapiro M, Slingo J, Molteni F. 2007. Collaborative research at the intersection of weather and climate. *World Meteorological Organization Bulletin* **56**(3): 1–9.
- Nakazawa T. 1988. Tropical super clusters with intraseasonal variations over western Pacific. *J. Meteor. Soc. Japan* **66**(6): 823–839.
- Neelin J, Zeng N. 2000. A quasi-equilibrium tropical circulation model-formulation. *Journal of the Atmospheric Sciences* **57**: 1741–1766.
- Pauluis O, Frierson DMW, Majda AJ. 2008. Precipitation fronts and the reflection and transmission of tropical disturbances. *Quarterly Journal of the Royal Meteorological Society* **134**(633): 913–930, doi: 10.1002/qj.250.
- Penland C. 1989. Random forcing and forecasting using principal oscillation pattern analysis. *Monthly Weather Review* **117**(10): 2165–2185.
- Stechmann S, Majda A. 2006. The structure of precipitation fronts for finite relaxation time. *Theoretical and Computational Fluid Dynamics* **20**(5-6): 377–404.
- Stechmann SN, Majda AJ. 2009. Gravity waves in shear and implications for organized convection. *J. Atmos. Sci.* **66**: 2579–2599.
- Straub K, Kiladis G. 2002. Observations of a convectively-coupled Kelvin waves in the eastern Pacific ITCZ. *J. Atmos. Sci.* **59**: 30–53.
- van Leeuwen P. 2010. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society* **136**: 1991–1999.
- Wheeler M, Kiladis G. 1999. Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber-frequency domain. *Journal of the Atmospheric Sciences* **56**: 374–399.
- Wheeler M, Kiladis GN, Webster PJ. 2000. Large-scale dynamical fields associated with convectively coupled equatorial waves. *J. Atmos. Sci.* **57**(5): 613–640.
- Wilks D. 2005. Effects of stochastic parameterizations in the Lorenz 96 model. *Quart. J. Roy. Meteor. Soc.* **131**: 389–407.
- Zagar N, Gustafsson N, Kallen E. 2004a. Dynamical response of equatorial waves in four-dimensional variational data assimilation. *Tellus A* **56**: 29–46.
- Zagar N, Gustafsson N, Kallen E. 2004b. Variational data assimilation in the tropics: The impact of a background-error constraint. *Quarterly Journal of the Royal Meteorological Society* **130**(596): 103–125, doi: 10.1256/qj.03.13.
- Zhang C. 2005. Madden-Julian oscillation. *Rev. Geophys.* **43**: RG2003, doi:10.1029/2004RG000158.

doi: 10.1175/JAS-D-11-053.1

Appeared Dec. 2011. 

for actual journal ref. A,  
which should be added.