Check for updates

Accepted Article

# Comparison of Conventional and Constrained Variational Methods for Computing Large-Scale Budgets and Forcing Fields

Paul E. Ciesielski*, Richard H. Johnson,
Department of Atmospheric Science,
Colorado State University,
Fort Collins, Colorado

Shuaiqi Tang,
Pacific Northwest National Laboratory, Richland, Washington

Yunyan Zhang, and Shaocheng Xie
Lawrence Livermore National Laboratory, Livermore, California

April 2021
Submitted to the Journal of Geophysical Research: Atmospheres

Key points:

- Merits and limitations of two methods for computing large-scale atmospheric budgets are examined
- Constrained variational analysis (CVA) method effectively minimizes many of the shortcomings of the conventional budget method
- However, accurate large-scale diagnostics in CVA are dependent on reliable background fields and rainfall constraints

*Corresponding author address: Paul E. Ciesielski, Department of Atmospheric Science, 1371 Campus Delivery, Colorado State University, Fort Collins, CO 80523

Email: paulc@atmos.colostate.edu

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2021JD035183.

This article is protected by copyright. All rights reserved.

Accepted Article

**Abstract**

Analyses of atmospheric heat and moisture budgets serve as an effective tool to study convective characteristics over a region and to provide large-scale forcing fields for various modeling applications. This paper examines two popular methods for computing large-scale atmospheric budgets: the conventional budget method (CBM) using objectively gridded analyses based primarily on radiosonde data and the constrained variational analysis (CVA) approach which supplements vertical profiles of atmospheric fields with measurements at the top of the atmosphere and at the surface to conserve mass, water, energy, and momentum. Successful budget computations are dependent on accurate sampling and analyses of the thermodynamic state of the atmosphere and the divergence field associated with convection and the large-scale circulation that influences it. Utilizing analyses generated from data taken during Dynamics of the Madden-Julian Oscillation (DYNAMO) field campaign conducted over the central Indian Ocean from October to December 2011, we evaluate the merits of these budget approaches and examine their limitations. While many of the shortcomings of the CBM, in particular effects of sampling errors in sounding data, are effectively minimized with CVA, accurate large-scale diagnostics in CVA are dependent on reliable background fields and rainfall constraints. For the DYNAMO analyses examined, the operational model fields used as the CVA background state provided wind fields that accurately resolved the vertical structure of convection in the vicinity of Gan Island. However, biases in the model thermodynamic fields were somewhat amplified in CVA resulting in a convective environment much weaker than observed.

This article is protected by copyright. All rights reserved.

Accepted Article

## 1. Introduction

Since the pioneering work of Yanai et al. (1973), heat and moisture budget analyses computed from an array of radiosonde stations have been used to examine the effects of convection on large-scale atmospheric motions. These analyses also provide large-scale forcing fields which are used by cloud-resolving models (CRMs) and for the study of parameterization schemes in single-column models (SCMs). In this approach, thermodynamic and wind fields from three or more upper-air sounding sites are interpolated onto a domain using an objective interpolation procedure (e.g., line integrals, least squares fitting, optimal interpolation). A thorough review of the commonly used methods for computing these budgets via interpolation of data is given in Zhang et al. (2001). The most popular and useful of these methods is referred to in this paper as the conventional budget method (CBM). One approach with this method is to first interpolate data onto a regular grid (e.g., equally spaced in latitude and longitude) and then use standard finite differences on the gridded fields to compute the terms in the budget equations.

The CBM is based on an approximate form of moist thermodynamics which begins with the assumption that moist static energy is materially conserved (except for radiative effects). Although the effects of approximate moist thermodynamics are generally small at a given time, under certain conditions they can be non-negligible. For example, Johnson (1980) estimated that the neglect of cloud storage effects resulted in errors on the order 20% in the column integrated moisture budget during periods of rapidly evolving cloud fields in the GARP Atlantic Tropical Experiment (GATE). Despite the limitations of using an approximate form of moist thermodynamics, the CBM formulation is well suited for use with radiosonde observations which provide all the needed fields.

Some noteworthy strengths of the CBM, which has been extensively used in past field campaigns, include the option to readily combine radiosonde data with thermodynamic and wind measurements from a variety of data platforms (e.g., satellite-based thermodynamic and winds retrievals, profiler winds). The general setup is such that derived fields are independent of models and their parameterization schemes, although model data on occasion are used in data-sparce

This article is protected by copyright. All rights reserved.

Accepted Article

regions to supplement the observations (Johnson et al. 2015). Limitations of this method, which have been discussed in previous literature, include the sensitivity of results to the particular interpolation scheme used and the scheme's parameters (Zhang et al. 2001), random sampling errors (Mapes et al. 2003), measurement uncertainty (Ciesielski et al. 2003), and missing data (Johnson et al. 2000).

Yet another limitation of the CBM is the sensitivity of its results to the configuration of the sounding network. For example, Ciesielski et al. (1999) found that the passage of an idealized Rossby wave over the triangular sounding array in the Atlantic Stratocumulus Transition Experiment (ASTEX) would result in a 20% error in divergence. A similar conclusion was reached in the study by Katsumata et al. (2011), which found large discrepancies between budget-derived and satellite estimates of rainfall during the Madden-Julian Oscillation (MJO) active phase (Madden and Julian 1972) in the 2006 Mirai Indian Ocean Cruise for the Study of the MJO Onset (MISMO) experiment (Yoneyama et al. 2008). Using numerically simulated MJO circulations they examined how different network configurations captured the divergent signal associated with the MJO. Their results showed that a quadrilateral array was superior to MISMO's triangular network at capturing the divergence signal associated with the Rossby and/or inertia–gravity wave components of the circulation. Since accurate sampling and analyses of the divergence field is a key to successful CBM computations, design of the sounding network should be given special consideration.

To overcome many of the limitations of the CBM, a constrained variational analysis (CVA) approach has been developed (Zhang and Lin 1997), which has subsequently become a popular alternative. In this approach the atmospheric state variables, typically derived from sounding data, are adjusted in a minimal way comparable to measurement uncertainty to conserve vertical constraints of mass, moisture, static energy, and momentum. In so doing, CVA produces diagnostic fields that are consistent with these constraints. As shown in Ghan et al. (2000), the dynamically and thermodynamically consistent large-scale forcing derived from

This article is protected by copyright. All rights reserved.

Accepted Article

CVA is critical to drive single-column and cloud-resolving models in testing physical parameterizations in weather and climate models. While the effects of measurement and sampling errors in sounding data are effectively minimized with CVA, accurate large-scale diagnostics in this method are dependent on reliable constraints, particularly rainfall (Xie et al. 2004) and, as will be shown in this paper, accurate background fields used to initialize the computations.

While satisfying these vertical constraints typically results in significant improvements in budget diagnostics, especially during periods when the original sounding data are missing or of questionable quality (Schumacher et al. 2007), use of model analyses as a background field may affect the usefulness of the CVA fields. This is particularly true if the model analyses contain biases resulting from their parameterization schemes. In fact, model analyses are often used in place of rawinsonde data when CVA is applied over small domains where radar rainfall is available but radiosonde coverage may be sparse. In addition, the need for long-term large-scale forcing datasets to drive single-column and cloud-resolving models to develop and test physical parameterization schemes in climate models has led to a variation of the CVA approach wherein operational model analyses are adjusted to balance column-averaged constraints (Xie et al. 2004).

Considering the wide use of the conventional and constrained variational approaches for computing large-scale budgets and forcing fields, this paper provides a detailed comparison of these methods using analyses from the Dynamics of the Madden-Julian Oscillation (DYNAMO) field campaign. Description of this experiment, its data, and application of CBM and CVA to these datasets is described in section 2. In section 3, we highlight some noteworthy applications of the budget methods and limitations of the CBM. The sensitivity of the CVA to background states and rainfall constraints are considered in section 4. The merits and limitations of each approach are summarized in section 5 with concluding remarks provided in Section 6.

## 2. Data and description of budget methods

*a. Data*

This article is protected by copyright. All rights reserved.

Accepted Article

The data used in this study come from the DYNAMO field campaign, conducted from October 2011 through March 2012 over the central Indian Ocean. This experiment was designed to examine a number of outstanding issues related to the intraseasonal oscillation (Yoneyama et al. 2013; Zhang 2013; Zhang et al. 2013). Its sounding array was composed of two quadrilateral arrays – one north and one south of the equator. This study will focus on the period from 2 October through 31 December 2011 (hereafter the OND period) when the overall sounding network was most complete with 4-8 sounding observations per day (Ciesielski et al. 2014) and MJO activity was well established (Gottschalck et al. 2013, Johnson and Ciesielski 2013). A supersite located at Gan Island (0.69°S, 73.2°E), part of the Atmospheric Radiation Measurement Program (ARM) MJO Investigation Experiment (AMIE), had multiple radars, and radiometers in addition to the 8 per day sounding observations. The National Center for Atmospheric Research S-PolKa (SPOL) radar was also located on Gan Island. In this study CBM-based rainfall rate estimates will be compared to those determined from the SPOL radar. The sounding and radar datasets were quality-controlled and bias corrected, as needed, as part of a special effort to create legacy datasets for this experiment (Ciesielski et al. 2014, Xu et al. 2015, and B. Dolan personal communication, 2017).

Other data used in this study include fractional cloud amount and optical depth provided by the Clouds and the Earth's Radiant Energy System (CERES) product at 3-h resolution on a 1° grid (Wielicki et al. 1996). Gan radiative heating profiles at 1-h, 25 hPa resolution were from the Combined Retrieval (CombRET) product (Feng et al. 2014). Rainfall data from the TRMM 3B42 V7 product at 0.25°, 3-h resolution (Huffman et al. 2007) were also used. Estimates of latent and sensible heat come from the TropFlux product (Praveen Kumar et al. 2012) which are available at daily resolution on a 1° grid. Model fields used in this study were from the European Centre for Medium-Range Weather Forecasts (ECMWF) operational analysis (OA) at 0.25° horizontal resolution, 20 vertical levels from the surface to 20 hPa, and 6-h intervals. These analyses were available for the OND period, which covers the period of focus for this study. A majority (~95%) of the data from the sounding

This article is protected by copyright. All rights reserved.

Accepted Article

arrays were transmitted to operational centers in real time, so that the ECMWF OA in the core sounding domain is heavily influenced by the sounding data.

To focus on the intraseasonal aspects of the DYNAMO analyses, an MJO composite analysis was created in which data were subjected to a low-pass (LP) Kaiser filter (Hamming 1989) in time to retain variability at frequencies 20 days and longer. With application of this LP filter, 6 days of 3-h data are lost at the ends of the filtered time series. Further details of the compositing technique can be found in Ciesielski et al. (2017). For the OND period examined in this study, the DYNAMO network was fortunate to capture three MJO events (Gottschalck et al. 2013), as will be seen in subsequent analyses.

*b. Conventional Budget Method (CBM)*

While a variety of approaches can be used within CBM framework, this section describes details of an approach that has been developed and refined over the past 30 years beginning with analysis for the Tropical Ocean Global Atmosphere Coupled Ocean–Atmosphere Response Experiment (TOGA COARE) conducted in 1992-93.

In constructing the CBM gridded analyses for this study, ECMWF OA was used at 5° grid intersections if no observations (soundings, satellite winds, or otherwise) were present within a 4.5° radius of such an intersection. To facilitate its use at 3-h intervals, the ECMWF OA was linearly interpolated in time to create a 3-h-resolution product. Since this procedure to enhance data coverage was only applied outside the main core arrays or when the ships were offsite, results in the interior are largely independent of model analyses and thus parameterizations of physical processes (Johnson et al 2015).

Following interpolation to 3-h intervals, the sounding data, along with the other observations (including dropsondes, ASCAT and satellite winds, COSMIC thermodynamic data, see Johnson et al. 2015 for details), and model fields described above, were objectively analyzed onto a 1° grid at the surface and at 25-hPa intervals from 1000 to 50 hPa over the entire large-scale sounding array (LSA) shown in Fig. 1. While many objective interpolation schemes are available for such

This article is protected by copyright. All rights reserved.

Accepted Article

purposes, a multiquadric (MQD) interpolation procedure has been chosen as described in Ciesielski et al. (1997), since this method has been shown to produce more reasonable analyses over other interpolation procedures (e.g., Barnes 1964, Cressman 1957) in sparce data regions (Nuss and Titley 1994). This was readily apparent during periods when ship soundings were missing in TOGA COARE as budget-derived rainfall estimates computed from MQD analyzed fields showed the most favorable comparison to independent satellite estimates (Johnson et al. 2000).

After initial analysis of the wind to a regular grid, two adjustments to the winds are made to improve the quality of the divergence field. First, because interpolation of irregularly spaced winds to a regular grid can result in spurious divergence, an adjustment procedure was applied to the horizontal winds to minimize this erroneous divergence. An example of this issue and a procedure to correct it are described in Haertel (2002). Next, the mass continuity equation is used to compute vertical motion $\omega$. Integration starts at the surface where $\omega_s = 0$ if over the ocean or a flat surface, and $\omega_s = -g\,\rho_s\,(u_s\,\frac{\partial h}{\partial x} + v_s\frac{\partial h}{\partial y})$ if over sloping terrain, where $u_s$ and $v_s$ are the surface wind components, $\rho_s$ is surface density, and $h$ is topographic elevation. The divergence field is then mass balanced in the vertical with a method suggested in O'Brien (1970) by assuming adiabatic vertical motion $\omega_a$ at the lapse-rate tropopause (LRT) level. The LRT is determined from the 25-hPa resolution gridded dataset at each horizontal grid point and time step using the WMO thermal tropopause definition (i.e., the lowest level at which the rate of decrease of temperature with respect to height decreases to 2 K km$^{-1}$, and the average from this level to any level within the next 2 km does not exceed 2 K km$^{-1}$). This more precise estimate of tropopause vertical motion (i.e., $\omega_a$ as opposed to $\omega = 0$) involves solving the thermodynamic equation for $\omega$ at the tropopause level where radiative and convective diabatic effects are assumed to be zero. As seen in Johnson et al. 2015 (Fig. 10), net radiation at Gan is near zero at the tropopause level, at least in a daily-mean sense. After mass balancing each vertical column, a second adjustment to the horizontal wind field is applied from the surface to the LTR level following a

This article is protected by copyright. All rights reserved.

Accepted Article

procedure described in Johnson and Ciesielski (2002) to ensure consistency between the winds and the mass-balanced divergence profile.

The time series of the lapse-rate tropopause (LRT) height at Gan (Fig. 2) shows variability on multiple time scales including: seasonal, MJO, and numerous shorter-term fluctuations. The gradual rise in LRT height over this 3-month period reflects a seasonal change related to enhanced tropical upwelling in boreal winter by the Brewer-Dobson circulation (e.g., Yulaeva et al. 1994). On the MJO time scale (~30 days) the LRT height rises gradually over the course of several days, then quickly descends in response to cooling associated with gravity wave activity induced by MJO convection (Dagg 2015). Over this 3-month period the LRT height ranged between 140 and 75 hPa with a mean height of 102 hPa. Correctly assigning the upper limit for the integration of divergence to the LRT height improves the computation of vertical motion. Since lapse rates $\left(\frac{\partial\theta}{\partial p}\right)$ are large in the lower stratosphere, diagnosing the correct vertical motion in the tropopause layer is important to avoid computation errors in $Q_1$. Accounting for tropopause variability becomes even more imperative in mid-latitude budget analyses where the passage of synoptic-scale systems can cause the LRT height to fluctuate by a few hundred hPa on short time scales.

Using centered finite differences with the gridded data, $Q_1$ (apparent heating) and $Q_2$ (apparent drying) are computed from equations (1) and (2) following Yanai et al. (1973).

$$Q_1 = \frac{\partial s}{\partial t} + \vec{V}\cdot\nabla s + \omega\left(\frac{\partial s}{\partial p}\right) \qquad (1)$$

$$Q_2 = -L\left[\frac{\partial q}{\partial t} + \vec{V}\cdot\nabla q + \omega\left(\frac{\partial q}{\partial p}\right)\right] \qquad (2)$$

With estimates of surface latent ($LE$) and sensible heat ($S$), integral forms of budget equations are used to compute surface precipitation ($P$) and column-net radiation $\langle Q_R \rangle$.

This article is protected by copyright. All rights reserved.

$$< Q_1 > = LP + < Q_R > + S \qquad (3)$$

$$< Q_2 > = LP - LE \qquad (4)$$

$$< Q_1 > - < Q_2 > = < Q_R > + S + LE \qquad (5)$$

$$\text{where.} < > = \frac{1}{g} \int_{p_{TOA}}^{p_s} ( ) \partial p \qquad (6)$$

is the vertical integral from the top of atmosphere (TOA) pressure to the surface pressure $p_s$, $L$ *is* the latent heat of vaporization, $\vec{V}$ the horizontal wind vector, $s = c_p T + gz$ the dry static energy, $q$ the water vapor mixing ratio, $S$ the surface sensible heat flux, $E$ the surface evaporation rate, $g$ the gravitational constant, and $Q_R$ the net radiative heating rate. The variables in (1)-(5) and (7)-(10) below represent horizontally-averaged qualities. Storage and advection of hydrometeors as well as ice processes are neglected in this formulation. Comparing the budget-residual quantities $P$ and $< Q_R >$ to reliable independent estimates allows one to determine the overall reliability of the budget-derived analyses (Johnson et al. 2000).

c. *Constrained Variational Analysis (CVA) Approach*

As shown by Zhang and Lin (1997), the column-integrated constraints are satisfied with the vertical integration of the following large-scale governing equations for atmospheric mass, moisture, dry static energy, and momentum:

$$\langle \nabla \cdot \vec{\boldsymbol{V}} \rangle = - \frac{1}{g} \frac{dp_s}{dt} \qquad (7)$$

$$\frac{\partial \langle q \rangle}{\partial t} + \langle \nabla \cdot \vec{\boldsymbol{V}} q \rangle = E - P - \frac{\partial \langle q_l \rangle}{\partial t} + \frac{\omega_s \, q_s}{g} \qquad (8)$$

This article is protected by copyright. All rights reserved.

Accepted Article

$$\frac{\partial\langle s\rangle}{\partial t} + \langle\nabla\cdot\vec{V}s\rangle = \langle Q_R\rangle + LP + S + L\frac{\partial\langle q_l\rangle}{\partial t} + \frac{\omega_s\, s_s}{g} \quad (9)$$

$$\frac{\partial}{\partial t}\langle\vec{V}\rangle + \langle\nabla\cdot\vec{\boldsymbol{V}}\,\vec{\boldsymbol{V}}\rangle + f\,\vec{k}\times\langle\vec{V}\rangle + \nabla\langle\phi\rangle = \overrightarrow{\tau_s} \quad (10)$$

where the bracket $\langle * \rangle$ is given by (6), $q_l$ is cloud liquid water content, $\omega_s$, $q_s$, and $s_s$ the pressure vertical velocity, water vapor mixing ratio and dry static energy at the surface, respectively, $f$ the Coriolis parameter; $\vec{k}$ the unit vector in vertical direction; $\phi$ the geopotential, and $\overrightarrow{\tau_s}$ the surface wind stress. Ice processes and advection of cloud hydrometeors are neglected.

In the CVA method, the atmospheric variables ($\vec{V}$, $s$, $q$) are forced to satisfy (7) – (10) with minimum adjustments to the first guess (either from radiosonde or from NWP analyses). The final analysis product is derived by minimizing the cost function:

$$I(t) = \iiint_{p,x,y}\left[\alpha_u(u-u_0)^2 + \alpha_v(v-v_0)^2 + \alpha_q(q-q_0)^2 + \alpha_s(s-s_0)^2\right]dxdydp \quad (11)$$

With (7)--(10) as strong constraints, where *u, v, q, s* denote the final analysis data, $u_0, v_0, s_0$ denote the first guess, and $\alpha$ is the weighting function related to the error estimates in the initial first guess, which are set as 0.5 m/s for winds, 0.2 K for temperature, and 3% for water mixing ratio.

The CVA approach shares some similarity with 3DVAR data assimilation (DA) in that they both use the variational method (i.e., minimizing a cost function) to calculate a best estimate state. However, the purpose of CVA is to achieve a state that is self-consistent with various constraints. Unlike DA methods that involve a numerical weather prediction (NWP) model for assimilating observations into the system, CVA does not require a forecast model. NWP analyses only serve as the initial guess, which will be replaced or adjusted by soundings, depending on the distance between an observation station and an analysis grid point. When a sounding network is unavailable, NWP analyses serve as a proxy for observations as

This article is protected by copyright. All rights reserved.

Accepted Article

in the case of creating long-term continuous forcing data for cloud modeling studies (Xie et al. 2004).

Using the CVA as described above, large-scale fields, including apparent heating $Q_1$ and drying $Q_2$, were computed at 3-h, 25-hPa resolution for the OND period when 10-min SPOL rain rates and 3-h soundings were available at Gan. The CVA products used here and referred to in previous literature (e.g., Yu et al. 2018) as the AMIE-Gan analyses (AGA), are representative of a near circular region with a 150 km radius centered on Gan (i.e., the constrained variational domain or CVD shown in Fig. 1). Two primary CVA analyses were used in this study: one constrained by TRMM 3B42 rainfall averaged over the CVD, and a second by the legacy SPOL radar rainfall. The legacy radar data represent a special reprocessing of the DYNAMO radar datasets using the best-known processing assumptions (Powell et al. 2016). While CVA is designed to include other constraints, like column-net radiation and surface pressure, only its sensitivity to rainfall is considered in this study, as it is by far the strongest constraint (Zhang et al 2001, Xie et al. 2004).

In the absence of a sounding network on the scale of the radar coverage around Gan (see Fig. 1), the background atmospheric state for the AGA was defined by sampling the ECMWF OA at eight grid points in a circle i.e., the CVD over which SPOL radar rainfall maps were available. The approach of constraining a background atmospheric state based on model analyses, as opposed to actual sounding observations, is described in Xie et al. (2004). Confidence for using the ECMWF OA as a realistic background field comes from the fact that it assimilated observations from the enhanced DYNAMO sounding network (Fig. 1) in real time. Since model biases can impact the CVA, Fig. 3 shows the mean differences between ECMWF OA fields interpolated to the Gan location and the Gan sounding data. Differences are generally small, with the exception of a 1°C model cool bias below 900 hPa and a moisture bias above 300 hPa where model RHs exceed sonde values by more than 20% above 150 hPa. The low-level cool bias may be related to island heating effects in the sounding data which are not present in model analyses since they are more representative of open ocean conditions (Ciesielski and Johnson 2021). The upper-

This article is protected by copyright. All rights reserved.

Accepted Article

Accepted Article

level moist bias suggests excessive high cloudiness in the model which, through radiative feedbacks, likely explains the slight model warm bias at 200 hPa and cool bias at 100 hPa (Ackerman et al. 1988). Moist biases near and above the tropopause were also observed in mid-latitude ECMWF model analyses (Dyroff et al. 2015) which they attribute, in part, to issues with the numerical diffusion of moisture. Mean wind differences are less than 1 m/s but indicate weaker easterly shear with height in the model fields. The impact of these model biases is considered in section 4.

## 3. Applications of budget methods and limitations of CBM

Application of the CBM on a gridded field can provide useful information regarding horizontal gradients of diagnosed fields. This is evident in Fig. 4 of Johnson et al. (2015) which compares horizontal rain maps diagnosed from the $Q_2$ budget to that from the TRMM 3B42 product for the October-November 2011 period over the LSA of DYNAMO. Because of the good data coverage during this period, including soundings, satellite-based datasets, and ECMWF OA outside the core sounding array, these maps show excellent agreement both in terms of the magnitude and areal distribution of rainfall with much higher rain rates south of the equator. In addition, CBM gridded products were able to accurately capture the diagnostic properties of a cross-equatorial Hadley cell during the 1-14 October period characterized by strong rising motion within an ITCZ between 5°S and 10°S and subsidence with little rainfall over the northern DYNAMO array (Ciesielski et al. 2018).

While CVA was originally formulated to provide area-averaged analyses over near circular, flat domains (Zhang and Lin 1997, Xie et al. 2004), recent enhancements to the CVA procedure have been made to provide capabilities similar to the CBM as described above. For example, Tang and Zhang (2015) presented and implemented a CVA version capable of providing three-dimensional (3D) analyses. This version, referred to as 3DCVA, was tested for a case study of a midlatitude cyclone occurring over the ARM southern Great Plains site and applied to study the scale dependence of SCM simulated precipitation and clouds (Tang et al. 2017). Using a grid covering a 4.5° by 5° area with 0.5° horizontal resolution, the scheme

This article is protected by copyright. All rights reserved.

Accepted Article

was able to successfully capture the 3D structure of a cyclone's diabatic heating fields. In addition, by using a terrain-following sigma vertical coordinate, Tang et al. (2020) were able to apply CVA to regions with sloped terrain.

While good spatial data coverage allows budget approaches to capture large-scale horizontal gradients in diagnosed fields, averaging such fields over an extended time period or a large area is necessary in the CBM to reduce errors associated with sampling errors. Following the procedure in Mapes et al. (2003), unresolved variability in the soundings is computed by taking the difference between CBM analyzed fields averaged over a 6° x 6° box centered on a sounding site (i.e., representing resolved variability) and actual sounding data. This was done for the six core DYNAMO sounding sites. These differences form a "noise" dataset from which profiles are randomly sampled and added to individual soundings to create a perturbed-sounding 20-member ensemble gridded dataset. Using ensemble standard deviation as an estimate of sampling error as in Mapes et al. (2003), Fig. 4 shows how the errors in rainfall are reduced by averaging over time and space. The areas over which the CBM analyses were averaged to create this plot are shown in Fig. 1. Thus, when presenting time series of CBM diagnosed fields, some simple averaging in time (e.g., 3 or 5-day running means) is needed to minimize random sampling errors. To better visualize this, Fig. 5a shows a time series of 3-h rainfall over the Gan region with and without filtering in time. Without filtering (thin black line) budget-derived rainfall is extremely noisy with frequent occurrence of unrealistic negative values. With filtering in time (heavy black curve) the correlation of the budget-estimated rainfall to the SPOL estimate (heavy red curve) increases from 0.5 to 0.9 giving us confidence in the budget estimate. Practically speaking, though, unfiltered CBM estimates of rainfall have limited use on time-scales shorter than a few days. However, meaningful analyses on shorter time scales (e.g., diurnal) can be produced by averaging CBM fields over several days (e.g., the diurnal cycle of ITCZ convection for a 2-week period in DYNAMO as discussed in Ciesielski et al. 2018). The period with negative 5-day filtered CBM rainfall in early December (Fig. 5a) is related to hydrometer storage and advection effects which are neglected in the CBM formulation (Schubert et al. 2018). Evidence for hydrometer

This article is protected by copyright. All rights reserved.

Accepted Article

storage can be inferred from changes in CERES high cloud fraction in Fig. 5b. The best example of this effect can be seen during the November period when large changes in high cloud amount are related to systematic differences between CBM and SPOL rainfall estimates. Further evidence for hydrometer effects during DYNAMO is presented in next section.

To assess the reliability of CBM basic fields, Fig. 6 shows vertical profiles of the period-mean difference between the Gan sounding data and the CBM analyzed fields averaged over the CVD. The CBM temperature field exhibits little mean bias (< 0.2°C at any given level). Whereas the moisture biases are generally in the direction of the ECMWF OA biases (except near the surface) (cf. Figs. 3 and 6) and of a similar magnitude, the u-component biases are slightly larger than the model's biases. Small differences between analyses and sounding data can be expected from the MQD interpolation scheme used by the CBM, which attempts to fit the analysis fields to the sounding data in a smooth fashion. Since horizontal temperature gradients are weak in the deep tropics, the small bias in this CBM field seems reasonable. Biases in the other CBM fields are related to the influence of surrounding data (satellite-based products) and use of ECMWF OA in the far field. While the CBM adequately captures the large-scale structure of the environment, as will be seen in next section, the coarseness of the input data fields (Fig. 1) limits its ability to accurately capture smaller-scale horizontal wind gradients and thus the divergence associated with convective features local to the Gan region.

## 4. Impact of background state and rainfall constraints on CVA analyses

This section highlights some of the strengths and limitations of CVA by considering how its analyses are affected by the background state and rainfall constraints. The CVA analyses using TRMM and SPOL as rainfall constraints will be referred to as CVAT and CVAS, respectively.

### *a. Impact of background state on CVA convective environment*

Biases in the CVAS basic fields relative to the Gan sounding data are shown in Fig. 7. Note: biases in CVAT are virtually identical to those in CVAS (not shown). Generally, the model biases shown in Fig. 3 are somewhat amplified in the CVAS fields. This is most pronounced in the low-level dry bias and upper-level westerly

This article is protected by copyright. All rights reserved.

wind bias which are more than doubled in the CVAS. While not present in the ECMWF OA (Fig. 3), a large warm bias is seen in the CVAS upper-level temperature field. This warm bias occurs in both the constant and observed tropopause version of the CVAS and is related to a smoothing procedure applied to the upper-level (above 140 hPa) CVA temperature field designed to eliminate unrealistically large $Q_1$ values. When CVA forcing data are used to drive SCMs/CRMs unrealistically large advective warming or cooling, caused by imbalances between the horizonal and vertical advection terms, cannot be offset by the warming/cooling generated by model physical processes (mainly radiation at upper levels) leading to large warm/cold biases in model simulations. The smoothing is applied to reduce unrealistically strong advective tendencies so that models can better respond to the imposed forcing. Recently, Tang and Zhang (2015) have shown that this smoothing, and its related issues, can be avoided by applying a radiative constraint in the upper troposphere.

To further examine the CBM and CVAS biases, Figs. 8 and 9 show time-height plots of perturbation temperature (anomalies from OND mean) and zonal winds from the CBM and CVAS. In both analyses descending tilted features above 200 hPa can be seen, which are more pronounced in the CBM fields. For example, the descending easterlies are typically 6 m/s stronger while the perturbation temperature anomalies are ~2°C larger in the CBM compared to the CVAS. Similar analyses based solely on Gan sounding data (not shown), indicate that the CBM has accurately captured these descending tilted features, while presumably the CVAS smoothing procedure has reduced their magnitude. Kiladis et al. (2001) and Randel and Wu (2005) attributed these tilted descending structures as a gravity or Kelvin wave response to the MJO convective heat source envelope. The descending temperature anomalies have been shown to impact convection through stability changes - enhancing (suppressing) instability when upper-level cold (warm) anomalies are present (Virts and Wallace 2014; Del Genio et al. 2012; Dagg 2015). Likewise, the descending zonal wind anomalies affect the shear profile which in turn will impact the characteristics of convection (Yamada et al. 2010; Lane 2021).

This article is protected by copyright. All rights reserved.

Accepted Article

Accepted Article

The bottom panel of Fig. 8 shows the time-height difference between the CBM and CVAS temperature fields. Relative to the CBM field, the CVAS exhibits a consistent warm bias near 100 hPa (~4°C) and a near-surface cool bias (~2°C). These temperature biases, along with the low-level CVAS dry bias seen in Fig. 7, impact the convective environment. This is illustrated in Table 1, which lists various convective parameters computed using the Gan sounding data, as well as the CBM and CVA fields. As seen here, the parameters computed with the CBM closely resemble those from Gan soundings, while those computed with the CVA depict a much weaker convective environment. For example, the mean CAPE computed with CBM fields is 94% of the mean Gan soundings value, while in CVAS fields it is only 52%. Also seen in Table 1 (far-right column), the mean CAPE computed from the ECMWF OA averaged over the CVD is 69%. This notion of a weaker convective environment in the CVAS fields is consistent with the CVA mean values of the LFC and the equilibrium level which are noticeably higher and lower, respectively, than the sounding and CBM values. In summary, the biases in the background thermodynamic fields are inherited and somewhat amplified by the CVAS (Fig. 3), and as a tradeoff to obtain reliable large-scale forcing fields, its smoothing procedure applied at upper levels degrades the accuracy of thermodynamic and wind fields. These limitations, which may have a significant impact on convective simulations that use the CVA large-scale forcing dataset, are likely present to some degree in other CVA field campaign datasets. For example, in a similar analysis of the Mid-latitude Continental Convective Clouds Experiment (MC3E; Xie et al. 2014) CVA had upper-level warm biases ~2°C or about half the magnitude of that seen in DYNAMO along with a moderately suppressed convective environment (e.g., CVA mean CAPE was reduced about 30% from MC3E sounding values.)

To test the idea that using an improved background state in the CVA would improve its convective environment, instead of using the ECMWF OA, the CVA was produced using the CBM gridded product along with Gan sounding data to initialize its background state. This version is referred to as CVASG (i.e., CVA with SPOL rainfall constraint and Gan soundings). The convective parameters for CVASG are listed in Table 1. As expected, the convective parameters for CVASG compare more

This article is protected by copyright. All rights reserved.

Accepted Article

favorably with those computed from the Gan soundings. For example, mean CAPE for this version is 82% of the Gan soundings values. Further improvement is limited by the upper-level warm bias in the CVASG, which is still ~4°C (not shown), and the CVA's tendency to amplify background state biases. For example, the slight near-surface cool bias in the CBM (0.15°C seen in Fig. 6) is doubled in the CVASG (not shown).

*b. Impact of background state on CVA diagnosed profiles*

To examine the impact of the background state on the vertical structure of the diagnosed fields, we first consider the time series of SPOL echo-top frequency and echo area coverage in Fig. 10. Within the slower evolution of convection on the MJO time scale (~30 days), the echo-top frequency (ETF) displays a myriad of high-frequency variabilities, such as the 2-day modes during the October MJO active phase (Zuluaga and Houze 2013; Yu et al. 2018). MJO convective stages, as defined in Ciesielski et al. (2017), are indicated along bottom (SP – suppressed, BH – bottom heavy, DC - deep convection, SF – stratiform). These stages denote the predominant convective type over the Gan region for each period. For example, while short-lived suppressed conditions were present between two strong Kelvin wave events characterized with deep convection over Gan, which occurred near the beginning and end of the 7-day DC stage of the November MJO (Gottschalck et al. 2013), the overall predominant mode was deep convection. Ruppert and Johnson (2015) divided their analyses into somewhat similar MJO categories, however their definitions were based on large-scale conditions over the NSA (Fig. 1).

Using these MJO stage definitions, Fig. 11a shows the mean ETF as a function of pressure for each stage. These profiles suggest that each stage has a mixture of convective types with the deep convection (echo tops above 300 hPa) occurring to some degree during all stages. Deep convection during the SP stage is likely from mesoscale cellular patterns of convection (Ruppert and Johnson 2015), which promote mid-level moistening and assists the transition from shallow to deeper convective clouds.

To evaluate the diagnosed fields from the various analyses, Fig. 12 shows the vertical profiles of convective heating $Q_1$-$Q_R$ and drying $Q_2$ for the different MJO

This article is protected by copyright. All rights reserved.

Accepted Article

stages from the CVAT, CVAS, and CBM analyses. Here the $Q_R$ profiles representative of the Gan region come from the radiative CombRet product (Feng et al. 2014). In general, all three analyses show a transition from shallow to deep to stratiform profiles over the course of the MJO, although important details vary among the analyses. The vertical structures are quite similar between the two CVA products with magnitude shifts in the profiles at various levels related to satisfying different rainfall constraints. This behavior is consistent with sensitivity tests conducted in prior studies (Waliser et al. 2002; Xie et al. 2006, 2010) which show that magnitudes of derived fields rather than their vertical structures are primarily affected by constraint uncertainties. Mean rainfall for each analysis and convective stage is listed in bottom panels. During the SP and BH stages when TRMM underestimated rainfall due to its lack of sensitivity at lighter rain rates (Xu and Rutledge 2014), the CVAS profiles are shifted towards greater heating and drying values associated with higher SPOL rain rates. Conversely, during the SF stage when TRMM 3B42 product overestimates rainfall due to the abundance of cirrus clouds and its use of IR data, the CVAS profiles are shifted towards lower heating and drying rates consistent with the smaller SPOL rain rates. These differences highlight the importance of using the correct rainfall constraint to achieve realistic convective structures during the various MJO stages.

Because of the importance of diabatic heating to prevailing MJO theories, many studies have examined the radiative and convective characteristic of the MJO using composite analyses (e.g., Jiang et al. 2011; Ciesielski et al. 2017). This perspective is presented in Fig. 13 which shows a 3 MJO composite analysis of the $Q_1$ and $Q_2$ fields for the different budget products considered in this study. Also seen in the right-most panels of Fig. 13 are a composite analysis of the CERES cloud area and optical depth, and lastly precipitation for the three analysis products. Here day 0 identified in Figs. 5 and 13 represents the peak in the low-pass filtered SPOL rainfall with negative (positive) lag days being prior to (after) this rainfall peak. Each of the products captures a vertical tilt with time in the convection related to the transition of convection from shallow to mid-level to deep to stratiform. However, the tilt varies somewhat among the products related primarily to different rainfall

This article is protected by copyright. All rights reserved.

Accepted Article

constraints. For example, the CVAT product has the least vertical tilt, due to TRMM's underestimation of rainfall in the build-up stages of the MJO resulting in a lack of low-level heating. This is in contrast to the CBM product which has the largest rainfall (bottom-right panel of Fig. 13) and deepest heating during this period. Also, the systematic rainfall differences between CBM and SPOL seen in this panel provide evidence for hydrometer effects. In short, during periods when the hydrometer field (clouds and precipitation) is rapidly increasing (decreasing) as in the MJO build-up (decay) stage (cf. Fig. 5b and composite cloud area and optical depth analyses in Fig. 13), budget-derived rainfall will overestimate (underestimate) rainfall.

Since differences among the analysis products in Fig. 12 are most pronounced during the BH stage, Fig. 14 examines the diagnostic fields averaged over the three BH periods for these products. To provide an independent assessment of the reliability of these analyses, the mean ETF profiles for these BH periods are shown in Fig. 11b. From the ETF profiles, we note that the first two BH periods (BH1 and BH2, respectively) are generally similar showing some deep convection, a broad secondary peak around 750 hPa, and a predominate peak below 900 hPa. In contrast, BH3 has a prominent peak at 500 hPa and a smaller peak at low levels compared to the other periods. Comparing these profiles to the $Q_1$-$Q_R$ profiles in Fig. 14 suggests that CVAS best captures the convective structures for these 3 periods with some deep heating in all periods, a predominant low-level heating peak in BH1 and BH2, a prominent mid-level peak in BH3. On the other hand, CVAT profiles show little, if any, heating above 400 hPa, while the CBM profiles suggest too much deep convection in all BH periods and poorly resolve the mid-level peak during BH3. Consistent with these heating features, the CVAS $Q_2$ profiles have low-level drying peaks during BH1 and BH2 and a large mid-level drying peak in BH3. These drying peaks at low and mid-levels during the different BH periods explain the rather peculiar looking double peak structure in the mean BH $Q_2$ profile (Fig. 12).

The differences between the CBM and CVA convective profiles for the BH periods can be attributed, in part, to differences in the upper-level divergence fields. For all BH periods, the CVA profiles are characterized with strong upper-level

This article is protected by copyright. All rights reserved.

Accepted Article

convergence above 250 hPa, in contrast to the CBM profiles which primarily have upper-level divergence. The upper-level convergence during the SP (not shown) and BH periods, which is especially prominent during BH1, is likely associated with a Kelvin wave excited by the MJO convection (Seo and Kim 2003; Schubert and Masarik 2006) or as identified by Powell and Houze (2015), as a dry eastward propagating wavenumber 1 disturbance in the zonal wind field. As postulated in the latter study, as the large-scale subsidence associated with upper-level wind field lessens, deep convection becomes more prominent, ushering in the active phase of the MJO.

To better understand the divergence differences between the CVA and CBM, Fig. 15 compares the BH period divergences for these analyses to those computed directly from the ECMWF OA winds at different horizontal resolutions. In the top panels we note that the divergence computed from the ECMWF OA winds sampled at 5° resolution (EC5D) generally looks similar to the divergence from the CBM. The largest discrepancies occur during the BH3 period when the EC5D divergence profile is positive at all levels. This 9-day, CVD-averaged profile shows how coarse sampling of model analyses can lead to some physically unrealistic conditions (i.e, non-zero column-net divergence), but also how the CBM can correct such background states by requiring mass balancing of the vertical column. On the other hand, in the bottom panels, the divergence computed from the ECMWF OA winds sampled at its full 0.25° resolution (ECQD) looks nearly identical to the divergence from the CVAS. Thus, the vertical structures of the CVA divergence and diagnostic fields are primarily determined by the background ECMWF OA winds with minor adjustments to satisfy the vertical constraints. This illustrates the importance of having an accurate background wind field in order to capture the correct vertical structure of the diagnostic fields. Conversely, the CBM approach used in this study attempts to minimize model inputs, but in so doing sacrifices the ability to capture accurately the divergence field associated with the convection that the ECMWF OA resolved. Despite its deficiencies, use of the CBM generated fields with Gan sounding data as a background field for the CVA does provide a modest improvement in the

This article is protected by copyright. All rights reserved.

vertical structure of the diagnosed fields (not shown) in addition to the significant improvement in the convective environment as highlighted in Table 1.

## 5. Merits and limitations of approaches

The following lists summarize the primary merits and limitations of each approach. When appropriate, specific examples based on DYNAMO analyses presented in this study are mentioned.

*Merits common to both approaches:*

- When data coverage is adequate, analyses are largely independent of operational model fields and inherent model biases associated with physical parameterizations.
- With adequate data resolution, horizontal gradients of diagnosed fields can be resolved (e.g., Johnson et al. 2015, Tang and Zhang 2015).
- Can be applied over any topographic surface, i.e., flat as well as sloped terrain (Johnson et al. 2007, Tang et al. 2020).
- When radiative profiles are available, the vertical eddy flux of moist static energy (which provides a measure of the activity of cumulus convection) can be computed (Ciesielski et al. 1999, Johnson et al. 2015, Yu et al. 2018).
- Can be applied in variety of convective settings, including nonprecipitating cloud fields (e.g., ASTEX).

*Merits specific to CBM:*

- Procedure has valuable integral constraints, such that surface rainfall and column-net radiation, which are computed as budget residuals, can be evaluated by comparison to independent estimates of these quantities. (Correlations between 5-day filtered budget-derived rainfall and SPOL radar rainfall were 0.9 giving us confidence in the CBM estimates.)
- Differences between observed and CBM-derived rainfall can provide evidence for hydrometer storage and advection effects as seen in the DYNAMO MJO composite analyses.

*Merits specific to CVA:*

Accepted Article

This article is protected by copyright. All rights reserved.

Accepted Article

- CVA computes large-scale budgets by adjusting sounding data within their measurement and sampling errors. Although these measurement and sampling errors are not corrected in CVA, the errors in the derived large-scale budgets are significantly reduced through use of integral constraints from both TOA and surface observations, primarily rainfall (Zhang et al. 2001, Xie et al. 2004).
- Diagnostic fields are consistent with integral constraints, which is critical for accurately driving single-column and cloud-resolving models.
- Overcomes limitations in spatial resolution of input observations by use of operational model analyses with constraints from surface and top of the atmosphere.
- Effects of hydrometeor storage, i.e., the $\left({}^{\partial q_l}/_{\partial t}\right)$ term in (8) and (9), can be included when such measurements are available.

*Limitations common to both methods:*

- Sensitivity to model biases in background state if the upper-air sounding coverage is poor.
- Reduced quality of analyses when measurements of rapidly evolving hydrometer fields (clouds and precipitation) are missing or of limited quality, which is typically the case.

*Limitations specific to CBM:*

- Sensitivity to the interpolation scheme, measurement errors in sounding data, configuration of the sounding network, and periods with missing data. Note: missing data can be replaced with model fields, but this introduces model biases into the analysis.
- Sensitivity to sampling errors in sounding data. Averaging over time and space is needed to minimize the impact of random sampling errors.
- Since column-integrated heat and moisture budgets are not constrained with observations, this may cause spurious budget residuals that make it difficult to interpret single-column and cloud-resolving model simulations when the diagnosed budgets are used to force these models.

This article is protected by copyright. All rights reserved.

Accepted Article

*Limitations specific to CVA:*

- Biases in background field can be amplified through application of integral constraints. (Biases in ECMWF OA thermodynamic fields were amplified in AGA which resulted in a convective environment that was much weaker than observed.)
- Sensitivity to errors in integral constraints; diagnosed fields are only as good as constraints used. Reliable radar-based precipitation estimates are typically limited to small areas. (Biases in the TRMM rainfall product over the course of MJO convective evolution adversely affected CVD diagnosed profiles.)
- The current implementation of CVA includes smoothing of fields near tropopause to reduce unrealistically large horizontal advection, which adversely effects analyses at upper levels. Smoothing resulted in a warm temperature bias (~4°C) near 100 hPa in the AGA and a muted signal in the descending wind and temperature anomalies at upper levels associated with gravity waves excited by MJO convection. These issues will likely impact simulated convection when using CVA large-scale forcing datasets. Use of an additional radiative constraint at upper levels could avoid this issue due to smoothing (Tang and Zhang 2020).
- No constraint on how fields are adjusted in vertical such that reliable background winds are needed to preserve relevant vertical structures in diagnosed fields. (CVA preserved vertical structure of background divergence in the AGA). This issue could be partially addressed by applying an additional constraint from precipitation radar retrieved latent heating profiles (if available), such as those derived from TRMM (e.g., Schumacher et al. 2004), to preserve analyzed vertical heating structures during strong convective cases.
- Because advection of hydrometers is not considered, precipitation generated outside of CVA domain and advected in, or generated inside of CVA domain and advected out, can result in an inconsistency between diagnosed fields

This article is protected by copyright. All rights reserved.

Accepted Article

and precipitation falling within domain. For example, hydrometers may be advected into an array but not generated by forcing fields within that array. While likely not a significant issue over larger domains, this may be a greater problem in CVA applied to a single radar domain (150 km x 150 km) when mesoscale convective systems traverse the region.

## 6. Concluding remarks

This paper examines the merits and limitations of two popular methodologies for computing heat and moisture budgets applied to observations from the DYNAMO experiment: the conventional budget method (CBM) patterned after Yanai et al. (1973) and the constrained variational analysis (CVA) approach of Zhang and Lin (1997). The success of these methods is dependent on accurate sampling and analyses of the thermodynamic state of the atmosphere and the divergence field associated with convection and the large-scale atmospheric circulation that forces it. The CBM as applied here interpolates atmospheric state variables primarily from sounding data onto a regular latitude/longitude grid, then uses finite differences to compute budget terms. Using observations and operational model analyses, the constrained variational approach (CVA) adjusts state fields in a minimal way comparable to measurement uncertainty to conserve vertical constraints of mass, moisture, static energy, and momentum.

This study represents a benchmark comparison of the CBM and CVA approaches for diagnosing large-scale atmospheric budgets with the goal of providing guidance for applications of these procedures. As summarized in the previous section, both approaches have their strengths and limitations. Whereas the present application of the CBM strives to maintain the fidelity of the observations while limiting model influence, use of lower resolution data typical of field campaigns degrades the accuracy of diagnoses of subnet-scale processes. While CVA overcomes many of the limitations inherent in the CBM, its use of operational model analyses for background fields, particularly in the absence of adequate sounding networks, leaves it susceptible to model biases. Biases in background thermodynamic fields

This article is protected by copyright. All rights reserved.

Accepted Article

used by the CVA, as well as its smoothing procedure at upper levels, impact the convective environment, such that users of past CVA large-scale forcing datasets should be aware that these issues likely affected their convective simulations. Further improvements in budget analyses could be realized by addressing certain algorithm deficiencies but also by designing measurement strategies in future field campaigns to minimize or offset limitations of these procedures.

*Acknowledgments* The authors thank Emily Riley for her assistance in constructing the noise model for the random sampling error analysis. The CERES radiation data were obtained from: http://ceres.larc.nasa.gov/products.php?product=SYN1deg, the Gan radiative heating profiles from: https:/www/arm.gov/data/pi/71, the SPOL legacy radar data from: https://data.eol.ucar.edu/project/DLDP, the TRMM rainfall data from: https://disc.gsfc.nasa.gov/datasets/TRMM_3B42_7/summary, and the AMIE-Gan CVA from: http://www.arm.gov/news/data/post/28211. In addition, we thank Wayne Schubert, Minghua Zhang, and Brenda Dolan for many insightful discussions, Rick Taft for his help with many of the figures, and three anonymous reviewers for their constructive comments. This research was supported by the National Science Foundation (NSF) under grant AGS-1853633. Work at Lawrence Livermore National Laboratory (LLNL) was supported by the DOE ARM program and performed under the auspices of the U. S. Department of Energy by LLNL under contract No. DE-AC52-07NA27344. S. Tang was partially supported by the "Enabling Aerosol-cloud interactions at GLobal convection-permitting scalES (EAGLES)" project (74358) funded by the DOE, Office of Science, Office of Biological and Environmental Research, Earth System Model Development program. Pacific Northwest National Laboratory (PNNL) is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830.

This article is protected by copyright. All rights reserved.

Accepted Article

## References


Ackerman, T. P., Liou, K. N., Valero, F. P., and Pfister, L., 1988: Heating rates in tropical anvils. *J. Atmos. Sci*., **45** (10), 1606-1623.

Barnes, S. L. (1964), A technique for maximizing details in numerical map analysis. *J. Appl. Meteorol.*, **3**, 396– 409.

Ciesielski, P.E. W.H. Schubert, and R.H. Johnson, 1999: Large-scale heat and moisture budgets over ASTEX Region. *J. Atmos. Sci.*, **56**, 3241-3261.

Ciesielski, P. E., R. H. Johnson, P. T. Haertel and J. Wang, 2003: Corrected TOGA COARE sounding humidity data: Impact on diagnosed properties of convection and climate over the warm pool. *J. Climate*, **16**, 2370-2384.

Ciesielski, P. E. and Coauthors, 2014: Quality controlled upper-air sounding dataset for DYNAMO/CINDY/AMIE: Development and corrections. *J. Atmos. Oceanic Technol*., **31**, 741–764.

Ciesielski, P. E., R. H. Johnson, X. Jiang, Y. Zhang, and S. Xie, 2017: Relationships between radiation, clouds, and convection during DYNAMO. *J. Geophys. Res. Atmos.*, **122**, 2529-2548.

Ciesielski, P.E., R.H. Johnson, W.H. Schubert, J. H. Ruppert Jr., 2018: Diurnal cycle of the ITCZ in DYNAMO. *J. Climate*, **31**, 4543-4562.

Ciesielski, P. E. and R. H. Johnson: 2021: Small island effects in DYNAMO and their impact on large-scale budget analyses. *J. Appl. Meteor. Climatol.,* **60**, 577-594.

Cressman, P. G. (1957), An operational objective analysis system. *Mon. Wea. Rev*., **87**, 367–374. https://doi.org/10.1175/1520-0493(1959)087<0367:AOOAS>2.0.CO;2.

Dagg, E. L., 2015: Tropical tropopause layer variability associated with the Madden-Julian Oscillation during DYNAMO. M.S. thesis, Department of Atmospheric Science. Colorado State University, 84 pp.

Del Genio, A. D., Y. Chen, D. Kim, and M.-S. Yao, 2012: The MJO transition from shallow to deep convection in CloudSat/CALIPSO data and GISS GCM simulations. *J. Climate*, **25**, 3755-3770.


This article is protected by copyright. All rights reserved.

Accepted Article

Feng, Z., S. A. McFarlane, C. Schumacher, S. Ellis, J. M. Comstock, and N. Bharadwaj, 2014: Constructing a merged cloud-precipitation radar dataset for tropical convective clouds during the DYNAMO/AMIE experiment at Addu Atoll. *J. Atmos. Oceanic Technol*., **31**, 1021–1042.

Ghan S., and Coauthors, 2000: An intercomparison of single column model simulations of summertime midlatitude continental convection. *J. Geophys. Res.,* **105**, 2091-2121.

Gottschalck, J., P. E. Roundy, C. J. Schreck III, A. Vintzileos, and C. Zhang, 2013: Large-scale atmospheric and oceanic conditions during the 2011–12 DYNAMO field campaign. *Mon. Wea. Rev.,* **141**, 4173-4196.

Haertel, P., 2002: Spurious divergence within objective analyses with application to TOGA COARE heat and moisture budgets. *25th Conference on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., 4B2, https://ams.confex.com/ams/25HURR/webprogram/meeting.html

Hamming, R. W., 1989: *Digital Filters*, Dover, 284 pp.

Huffman, G. J., R. F. Adler, D. T. Bolvin, G. Gu, E. J. Nelkin, K. P. Bowman, E. F. Stocker, and D. B. Wolff, 2007: The TRMM multisatellite precipitation analysis: Quasi-global, multiyear, combined-sensor precipitation estimates at fine scale. *J. Hydrometeor*, **8**, 38–55.

Jiang, X., and Coauthors, 2011: Vertical diabatic heating structure of the MJO: Intercomparison between recent reanalyses and TRMM. *Mon. Wea. Rev*., **139**, 3208–3223.

Johnson, R. H., 1980: Diagnosis of convective and mesoscale motions during Phase III of GATE. *J. Atmos. Sci.,* **37,** 733–753.

Johnson, R. H., and P. E. Ciesielski, 2000: Rainfall and radiative heating rates from TOGA COARE atmospheric budgets. *J. Atmos. Sci.,* **57,** 1497–1514.

This article is protected by copyright. All rights reserved.

Accepted Article

Johnson, R. H., and P. E. Ciesielski, 2002: Characteristics of the 1998 summer monsoon onset over the Northern South China Sea. *J. Meteor. Soc. of Japan*, **80**, 561-578.

Johnson, R. H., P. E. Ciesielski, B. D. McNoldy, P. J. Rogers, and R. K. Taft, 2007: Multiscale variability of the flow during the North American Monsoon Experiment. *J. Climate*, **20**, 1628-1648.

Johnson, R. H., P. E. Ciesielski, T. S. L'Ecuyer, and A. J. Newman, 2010: Diurnal cycle of convection during the 2004 North American Monsoon Experiment. *J. Climate*, **23**, 1060-1078.

Johnson, R. H., and P. E. Ciesielski, 2013: Structure and properties of Madden–Julian oscillations deduced from DYNAMO sounding arrays. *J. Atmos. Sci.*, **70**, 3157–3179.

Johnson, R. H., P. E. Ciesielski, J. H. Ruppert Jr., and M. Katsumata, 2015: Sounding-based thermodynamic budgets for DYNAMO. *J. Atmos. Sci.*, **70**, 598-622.

Katsumata, M., P. E. Ciesielski, and R. H. Johnson, 2011: Evaluation of budget analysis during MISMO. *J. Appl. Meteor. Clim.*, **50**, 241-254.

Kiladis, G. N., K. H. Straub, G. C. Reid, and K. S. Gage, 2001: Aspects of interannual and intraseasonal variability of the tropopause and lower stratosphere. *Quart. J. Roy. Meteor. Soc.*, **127**, 1961–1984.

Lane, T. P., 2021: Does lower-stratospheric shear influence the mesoscale organization of convection? *Geo. Res. Let.*, ***48***, e2020GL091025.

Madden, R. A., and P. R. Julian, 1972: Description of global-scale circulation cells in the Tropics with a 40–50 day period. *J. Atmos. Sci.,* **29,** 1109– 1123.

Nuss, W. A., and D. W. Titley, 1994: Use of multiquadric interpolation for meteorological objective analysis. *Mon. Wea. Rev.,* **122,** 1611–1631.

Powell, S. W., and R. A. Houze Jr., 2015: Effect of dry large-scale vertical motions on initial MJO convective onset. *J. Geophys. Res. Atmos.,* **120**, 4783–4805, doi:10.1002/2014JD022961.

This article is protected by copyright. All rights reserved.

Accepted Article

Powell, S. W., R. A. Houze, Jr., and S. R. Brodzik, 2016: Rainfall-type categorization of radar echoes using polar coordinate reflectivity data. *J. Atmos. Oceanic Technol*., **33**, 523-538.

Praveen Kumar, B., J. Vialard, M. Lengaigne, V. S. N. Murty, and M. J. McPhaden, 2012: TropFlux: Air-sea fluxes for the global tropical oceans—Description and evaluation. *Climate Dyn*., **38**, 1521–1543, doi:10.1007/s00382-011-1115-0.

Randel, W. J., and F. Wu., 2005: Kelvin wave variability near the equatorial tropopause observed in GPS radio occultation measurements. *J. Geophys. Res*., **110**, D030102, doi:10.1029/2004JD005006.

Rowe, A. K., S. A. Rutledge, T. J. Lang, P. E. Ciesielski, and S. M. Saleeby, 2008: Elevation-dependent trends in precipitation observed during NAME. *Mon. Wea. Rev.*, **136**, 4962–4979.

Ruppert Jr, J. H., and R. H. Johnson, 2015: Diurnally modulated cumulus moistening in the preonset stage of the Madden–Julian oscillation during DYNAMO. *J. Atmos. Sci., **72***, 1622-1647.

Schubert, W. H., and M. T. Masarik, 2006: Potential vorticity aspects of the MJO. *Dynamics Atmos. and Oceans,* **42**, 127-151.

Schubert, W. H., P. E. Ciesielski, and R. H. Johnson, 2018: Heat and moisture budget analysis with an improved form of moist thermodynamics. *arXiv:1810.11119 [physics.ao-ph]*, 14 pp.

Schumacher, C., R. A. Houze Jr., and I. Kraucunus, 2004: The tropical dynamical response to latent heating estimates derived from the TRMM precipitation radar. *J. Atmos. Sci*., **61**, 1341-1358.

Schumacher, C., M. H. Zhang, and P. E. Ciesielski, 2007: Heating structures of the TRMM field campaigns. *J. Atmos. Sci*., **64**, 2593-2610.

Seo, K.-H., and K.-Y. Kim, 2003:, Propagation and initiation mechanisms of the Madden-Julian Oscillation. *J. Geophys. Res.*, **108** (D13), 4384, doi:10.1029/2002JD002876.

This article is protected by copyright. All rights reserved.

Accepted Article

Tang, S., and M. Zhang, 2015: Three-dimensional constrained variational analysis: Approach and application to analysis of atmospheric diabatic heating and derivative fields during an ARM SGP intensive observational period. *J. Geophys. Res. Atmos.*, **120,** 7283-7299**,** 10.1002/2015JD023621.

Tang, S., M. Zhang, and S. Xie, 2017: Investigating the scale dependence of SCM simulated precipitation and clouds by using gridded forcing data at SGP. *J. Geophys. Res. Atmos.*, **122**, 8724–8738, https://doi.org/10.1002/2017JD026565

Tang, S., S. Xie, M. Zhang, and S. Endo, 2020: Improvement of atmospheric objective analysis over sloping terrain and its impact on shallow-cumulus clouds in large-eddy simulations. *J. Geophys. Res. Atmos.,* **125**, e2020JD032492. https://doi.org/10.1029/2020JD032492.

Virts, K. S., and J. M. Wallace, 2014: Observations of temperature, wind, cirrus, and trace gases in the tropical tropopause transition layer during the MJO. *J. Atmos. Sci.*, **71**, 1143-1157.

Waliser, D.E., J. A. Ridout, S. C. Xie, and M. H. Zhang, 2002: Variational objective analysis for atmospheric field programs: a model assessment. *J. Atmos. Sci.*, **59**, 3436–3456.

Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee III, G. L. Smith, and J. E. Cooper, 1996: Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System experiment. *Bull. Amer. Meteor. Soc.*, **77**, 853–868.

Yamada, H., K. Yoneyama, M. Katsumata, and R. Shirooka, 2010: Observations of a super cloud cluster accompanied by synoptic-scale eastward-propagating precipitating systems over the Indian Ocean. *J. Atmos. Sci.,* **67**, 1456–1473, https://doi.org/10.1175/2009JAS3151.1.

Yanai, M., S. Esbensen, and J.-H. Chu, 1973: Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *J. Atmos. Sci.,* **30**, 611–627.

This article is protected by copyright. All rights reserved.

Yoneyama, K. C., and Coauthors, 2008: MISMO field experiment in the equatorial Indian Ocean. *Bull. Amer. Meteor. Soc*., **89**, 1889–1903.

Yoneyama, K., C. Zhang, and C. N. Long, 2013: Tracking pulses of the Madden–Julian oscillation. *Bull. Amer. Meteor. Soc*., **94**, 1871– 1891.

Yu, H., R. H. Johnson, P. E. Ciesielski, and H.-C. Kuo, 2018: Observation of quasi-2-day convective disturbances in the equatorial Indian Ocean during DYNAMO. *J. Atmos. Sci.*, **75**, 2867-2888.

Yulaeva, E., J. Holton, and J. M. Wallace, 1994: On the cause of the annual cycle in tropical lower-stratospheric temperatures. *J. Atmos. Sci*., **51**, 169-174.

Xie, S.C., R. T. Cederwall, and M. H. Zhang, 2004: Developing long-term single-column model/cloud system–resolving model forcing data using numerical weather prediction products constrained by surface and top of the atmosphere observations. *J. Geophy. Res. Atmos.,* **109**, D01104.

Xie, S.C., S. A. Klein, M. H. Zhang, J. J. Yio, R. T. Cederwall, and R. McCoy, 2006: Developing large-scale forcing data for single-column and cloud-resolving models from the Mixed-Phase Arctic Cloud Experiment. *J. Geophys. Res.*, **111**(D19), D19104.

Xie, S.C., T. Hume, C. Jakob, S. A. Klein, R. B. McCoy, and M. H. Zhang, 2010: Observed large-scale structures and diabatic heating and drying profiles during TWP-ICE. *J. Climate*, **23**, 57–79.

Xie, S.C., Y. Zhang, S. E. Giangrande, M. P. Jensen, R. McCoy, and M. H. Zhang, 2014: Interactions between cumulus convection and its environment as revealed by the MC3E sounding array. *J. Geophys. Res. Atmos*., **119**, 11,784–11,808, doi:10.1002/2014JD022011.

Xu, W. and S. A. Rutledge, 2014: Convective characteristics of the Madden–Julian oscillation over the central Indian Ocean observed by shipborne radar during DYNAMO. *J. Atmos. Sci.,* **71**, 2859-2877.

This article is protected by copyright. All rights reserved.

Accepted Article

Xu, W., S. A. Rutledge, C. Schumacher, M. Katsumata, 2015: Evolution, properties, and spatial variability of MJO convection near and off the equator during DYNAMO. *J. Atmos. Sci.*, **72**, 4126-4147.

Zhang, C., 2013: The Madden–Julian oscillation: Bridging weather and climate. *Bull. Amer. Meteor. Soc.*, **94**, 1849–1870.

Zhang, C. J. Gottschalck, E. D. Maloney, M. W. Moncrieff, F. Vitart, D. E. Waliser, B. Wang, and M. C. Wheeler, 2013: Cracking the MJO nut. *Geophys. Res. Let.,* **40**, 1223–1230, doi: 10.1002/ grl.50244.

Zhang, M. H., and J. L. Lin, 1997: Constrained variational analysis of sounding data based on column-integrated budgets of mass, heat, moisture, and momentum: Approach and application to ARM measurements. *J. Atmos. Sci.*, **54**, 1503-1524.

Zhang, M. H., J. L. Lin, R. T. Cederwall, J. J. Yio, and S. C. Xie, 2001: Objective analysis of ARM IOP data: Method and sensitivity. *Mon. Wea. Rev.* **129**, 295-311.

Accepted Article

This article is protected by copyright. All rights reserved.

Accepted Article

| parameter | Gan soundings | CBM | CVAS | CVASG | ECMWF OA |
|---|---|---|---|---|---|
| CAPE (J /kg) | 1294 | 1219 | 668 | 1065 | 889 |
| CIN (J / kg) | -54 | -49 | -51 | -50 | -35 |
| PW (mm) | 53.0 | 52.6 | 51.7 | 52.8 | 52.7 |
| LCL (hPa) | 936.7 | 937.9 | 946.4 | 943.0 | 946.4 |
| LFC (hPa) | 887.0 | 888.9 | 802.5 | 864.7 | 895.6 |
| Equilibrium Level (hPa) | 144.4 | 145.5 | 167.3 | 154.9 | 156.8 |

TABLE 1. Convective parameters for the OND period computed with Gan soundings CBM, CVA products, and ECMWF OA. CVAS refers to CVA with SPOL rainfall constraint and ECMWF OA used for background field, CVASG refers to CVA with SPOL rainfall constraint but with Gan soundings and CBM fields as background state.

This article is protected by copyright. All rights reserved.

Accepted Article

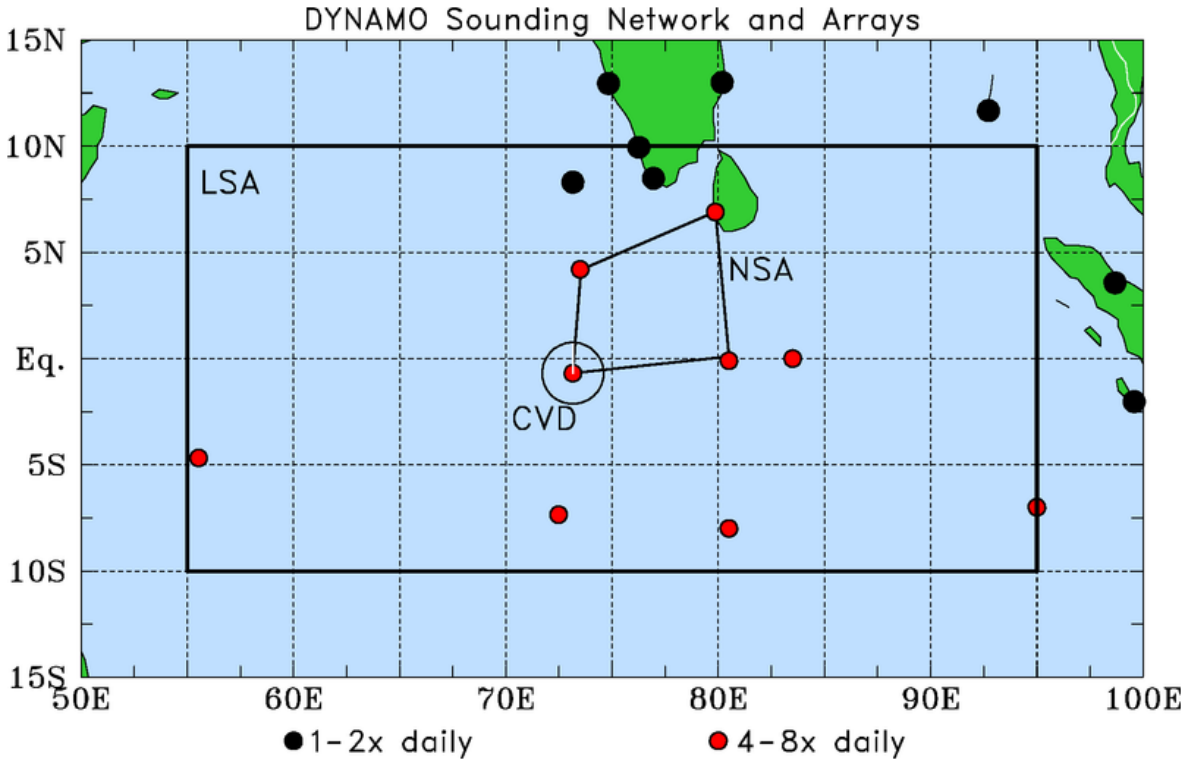


FIG. 1. Map of DYNAMO upper-air sounding network where color coding of dots denotes frequency of sonde launches at a site. Polygons show the averaging areas for the random sampling analyses which include the Large-scale Sounding Array (LSA) extending from 50° to 100°E, 10°N to 10°S, the Northern Sounding Array (NSA), and the Constrained Variational Domain (CVD) covering a circular area with a 150 km radius centered on Gan Island (73.15°E, 0.69°S).

This article is protected by copyright. All rights reserved.

Accepted Article

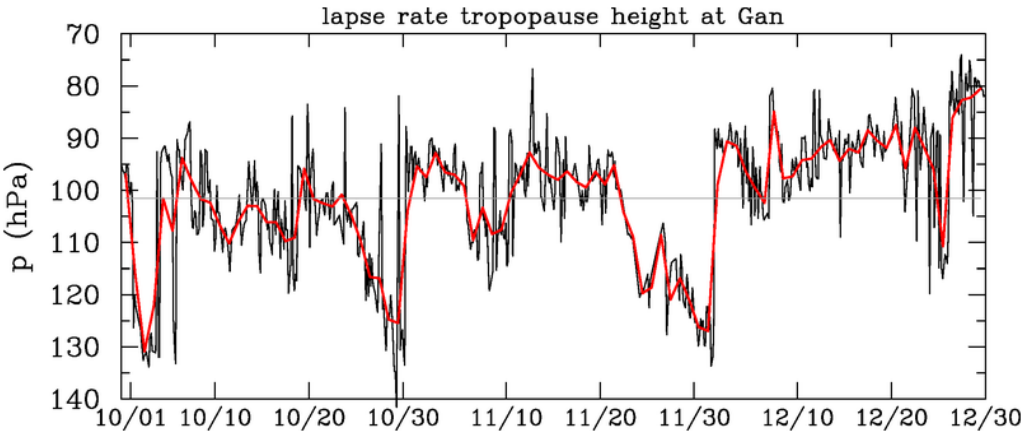


FIG. 2. Time series of lapse-rate tropopause height at Gan computed from 5hPa-resolution Gan sounding data. Black curve is based on 3-h soundings, red curve represents the daily-averages. Line gray line shows 3-month mean tropopause height which was at 102 hPa.

This article is protected by copyright. All rights reserved.

Accepted Article

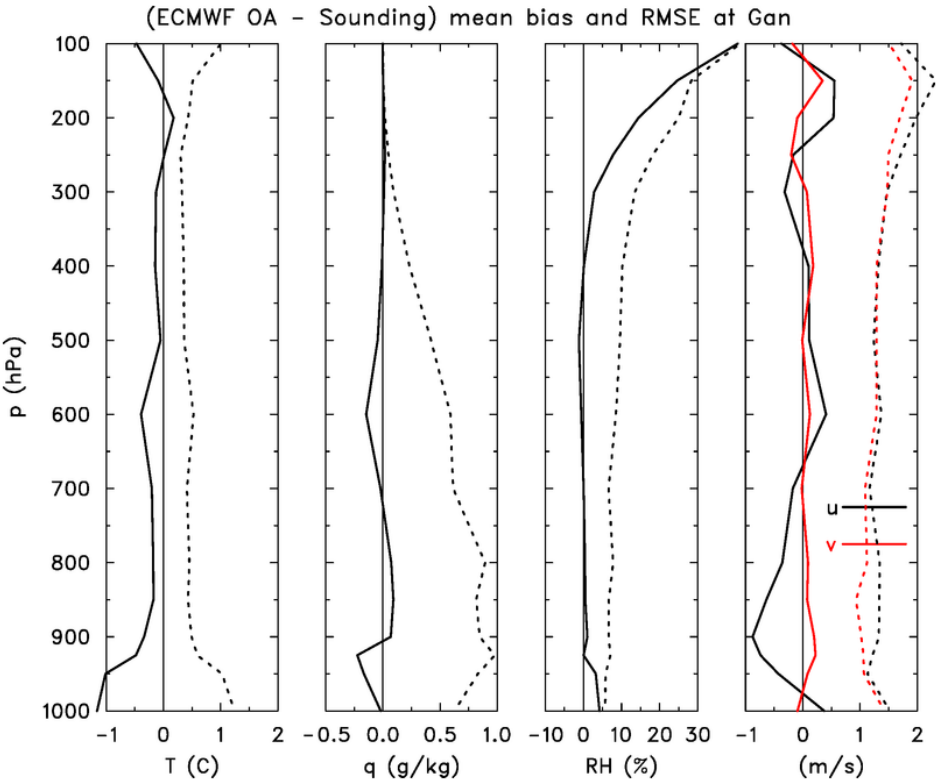


FIG. 3. OND-mean bias (solid) and RMSE (dashed) for ECMWF OA interpolated to Gan sounding site minus Gan sounding data fields for (from left to right): temperature, specific humidity, relative humidity with respect to ice, zonal wind (black) and meridional wind (red).

This article is protected by copyright. All rights reserved.

Accepted Article

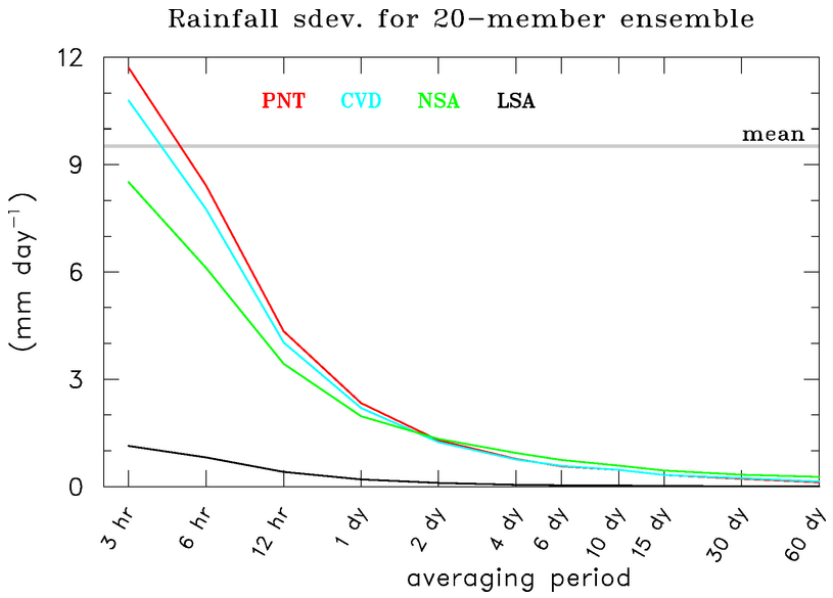


FIG. 4. Ensemble standard deviation of rainfall rate as a function averaging time, averaged over the indicated geographical areas shown in Fig. 1. PNT refers to the analyses at the grid point 73°E, 1°S. CVD is the CVA domain used in this study. Horizontal line represents the NSA mean for 3-month DYNAMO period.

This article is protected by copyright. All rights reserved.

Accepted Article

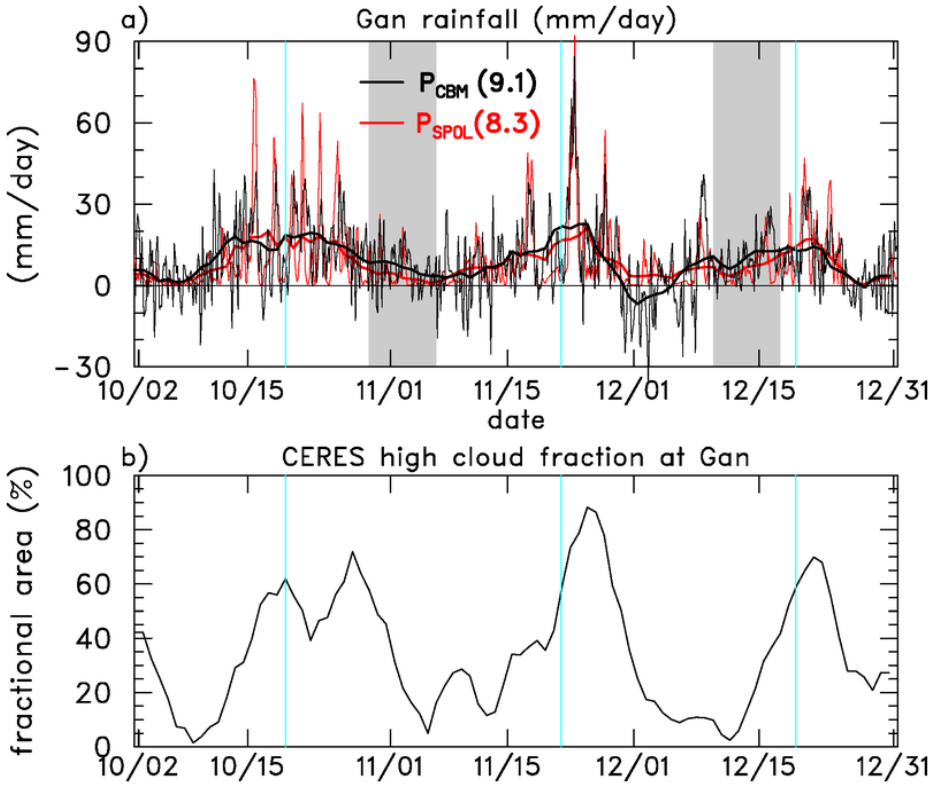


FIG. 5. a) Time series of rainfall at Gan averaged over CVD (i.e., the SPOL radar domain): 3-h CBM estimate (thin black), 5-day running mean filtered budget estimate (heavy black), 3-h SPOL estimate (thin red), and 5-day running mean filtered SPOL estimate (heavy red). Long-term period means listed in parentheses. Shading indicated periods when the Revelle was offsite and model influence may be larger on CBM analyses. b) Time series of 5-day running mean filtered CERES high cloud fraction averaged over a circle with a 1° radius centered at Gan. Thin cyan vertical lines indicate time of low-pass filtered SPOL rainfall peak for each MJO event (i.e. day 0 for composite analysis).

This article is protected by copyright. All rights reserved.

Accepted Article

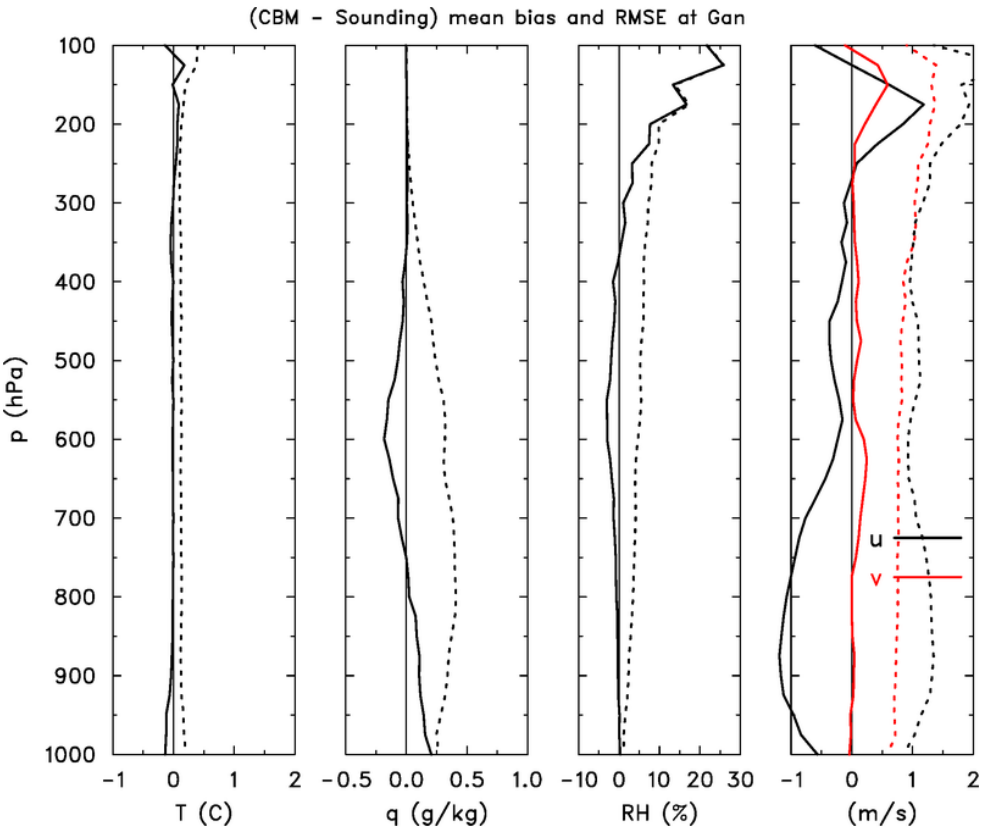


FIG. 6. As in Fig. 3, except OND-mean bias and RMSE for CBM fields minus Gan sounding data.

This article is protected by copyright. All rights reserved.

Accepted Article

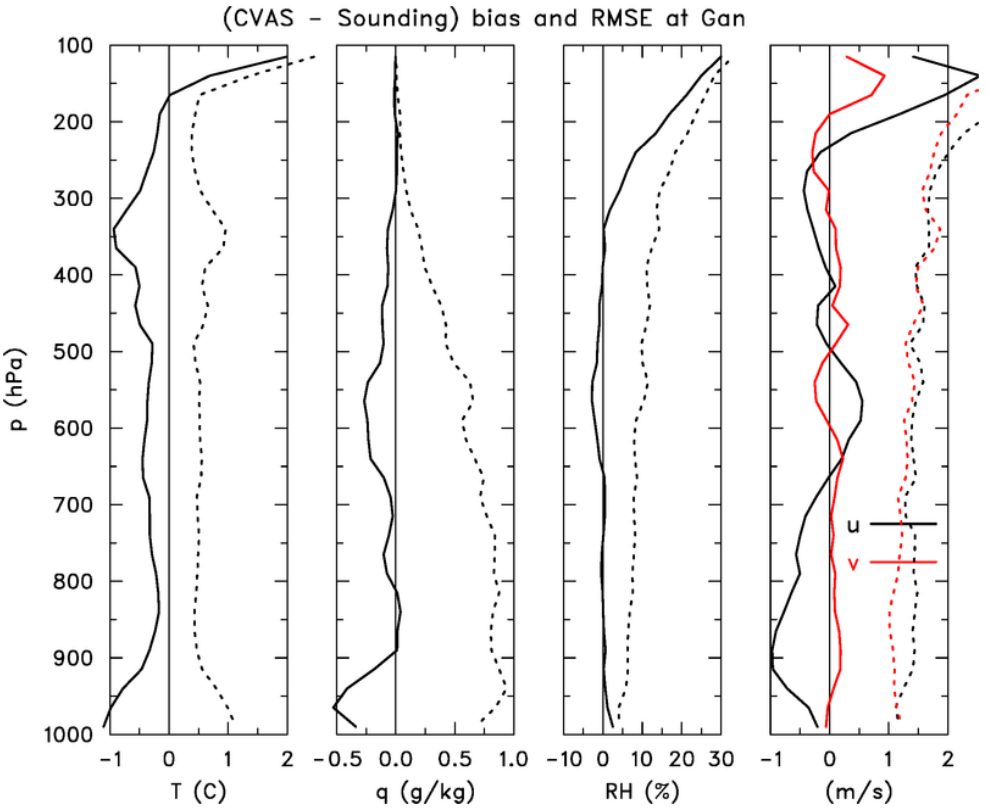


FIG. 7. As in Fig. 3, except for OND mean bias and RMSE for CVAS fields minus Gan sounding data.

This article is protected by copyright. All rights reserved.

Accepted Article

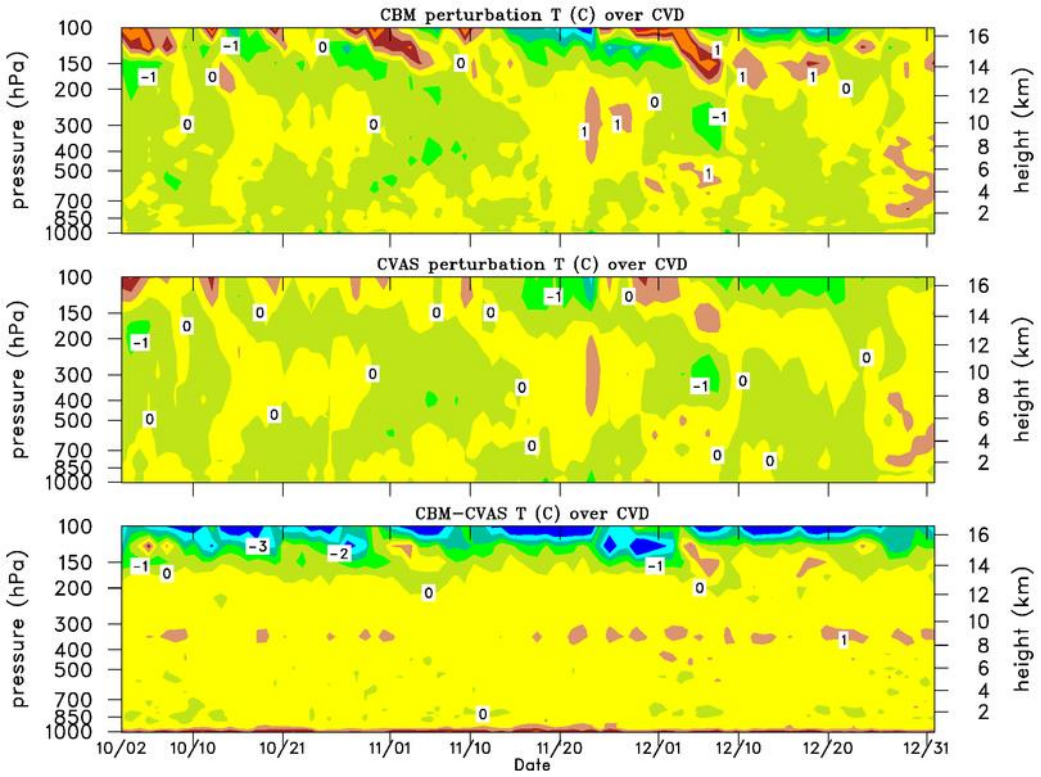


FIG. 8. (top panel) Time-height plot of perturbation temperature from CBM analysis over CVD (contour interval is 1°C) with cool (warm) shading indicating a negative (positive) perturbation from the period mean, (middle panel) as in top panel, except from CVAS analysis. (bottom panel) As above, except difference between CBM and CVAS temperature with cool (warm) shading indicating a negative (positive) difference.

This article is protected by copyright. All rights reserved.

Accepted Article

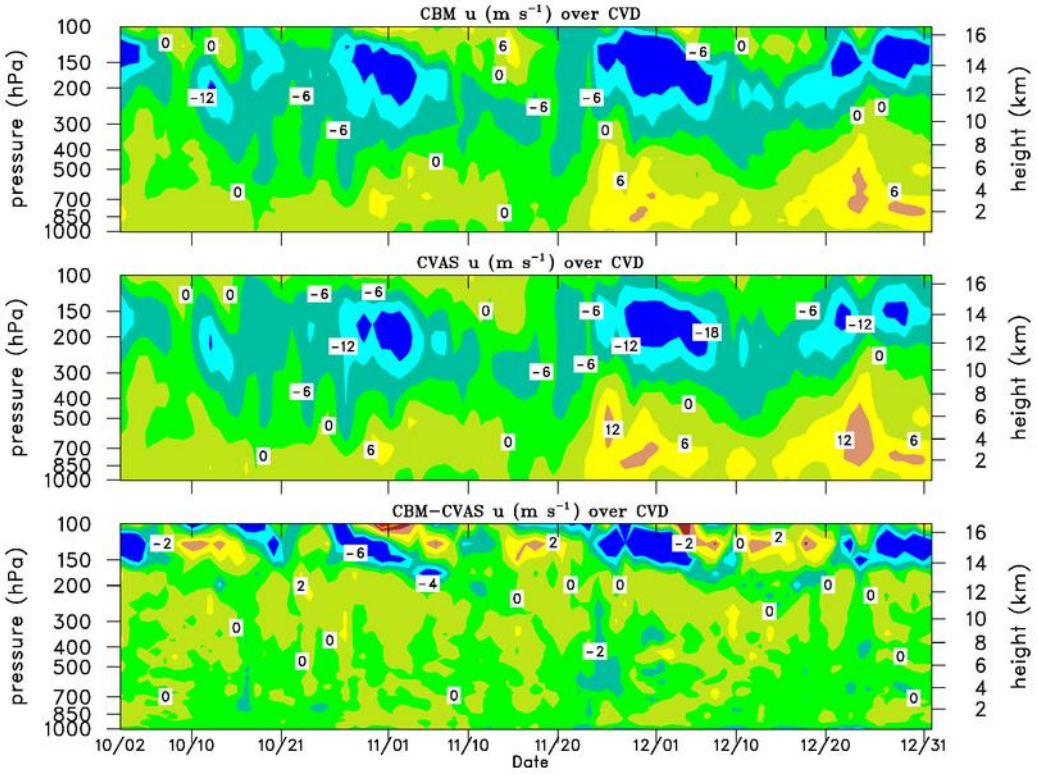


FIG. 9. (top panel) Time-height plot of zonal wind from CBM analysis over CVD (contour interval is 6 m/s) with cool (warm) shading indicating an easterly (westerly) wind component, (middle panel) as in top panel, except from CVAS analysis. (bottom panel) As above, except difference between CBM and CVAS zonal wind component with cool (warm) shading indicating a negative (positive) difference (contour interval is 2 m/s).

This article is protected by copyright. All rights reserved.

Accepted Article

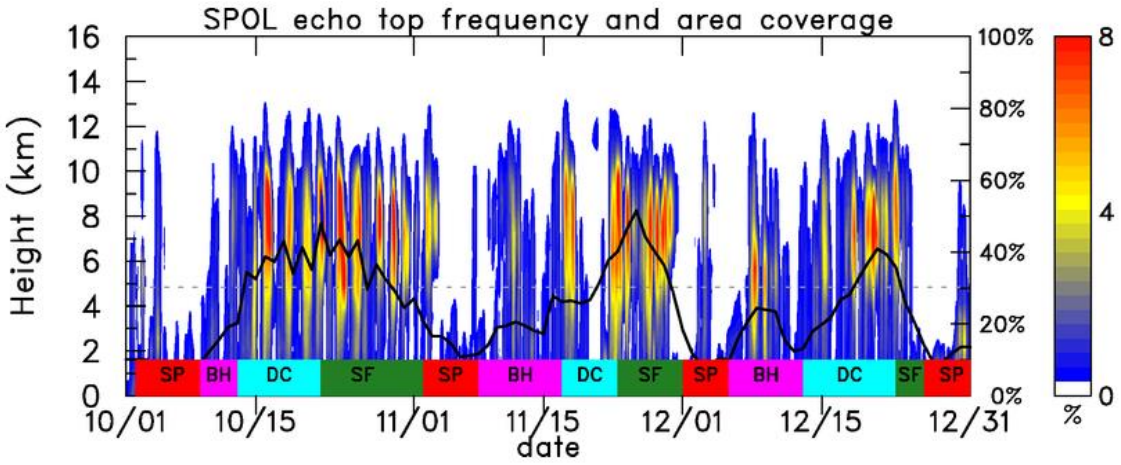


FIG. 10. Time series of echo area coverage (black curve in % with scale to right) and echo top frequency (shading; %) with scale bar to far right. A temporal smoothing was applied to echo area coverage (echo top frequency) using a 5 day (6 h) running mean. Missing data were filled in with linear interpolation in time before filtering was applied. Gray dashed line near 5 km represents period-mean 0°C level. Color bar along bottom of graph denotes various MJO stages, indicating the predominate type of convection in the vicinity of the Gan SPOL radar during that period as defined in Ciesielski et al. 2017; "SP" refers to suppressed, "BH" to bottom heavy, "DC" to deep convection, and "SF" to stratiform.

This article is protected by copyright. All rights reserved.

Accepted Article

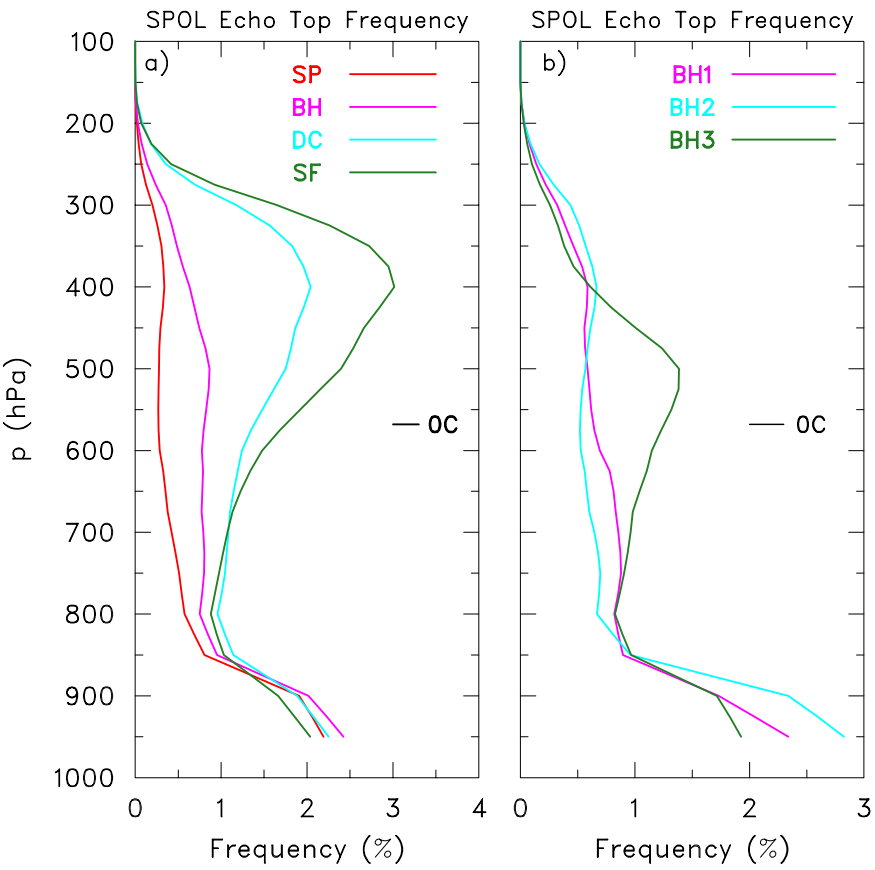


FIG.11. Mean echo top frequency (%) as a function of pressure based on SPOL radar data for (a) various MJO convective stages, and (b) for three BH periods.

This article is protected by copyright. All rights reserved.

Accepted Article

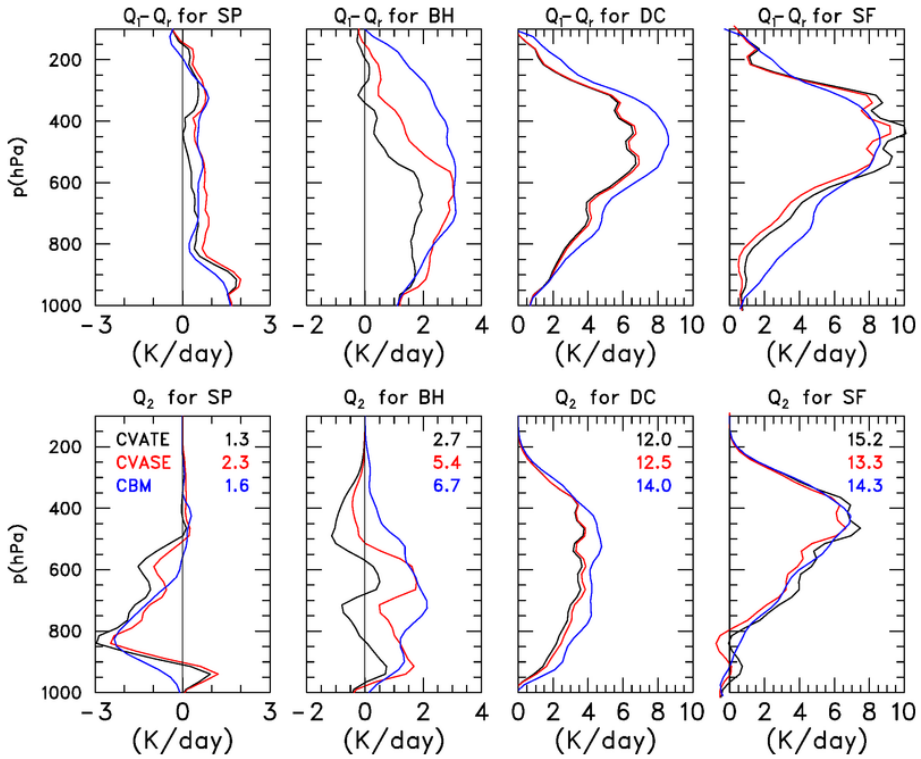


FIG. 12 Mean CVD vertical profiles from different analysis products (CVAT – black, CVAS – red, CBM – blue) for (top panels) apparent convective heating $Q_1$-$Q_R$ and (bottom panels) apparent drying $Q_2$ for the various MJO convective stages (left to right) SP, BH, DC and SF. Period mean rain rates (mm/day) are listed in bottom panels where CVAT is based on TRMM 3B42, CVAS is based on SPOL, and CBM is the $Q_2$ budget residual.

This article is protected by copyright. All rights reserved.

Accepted Article

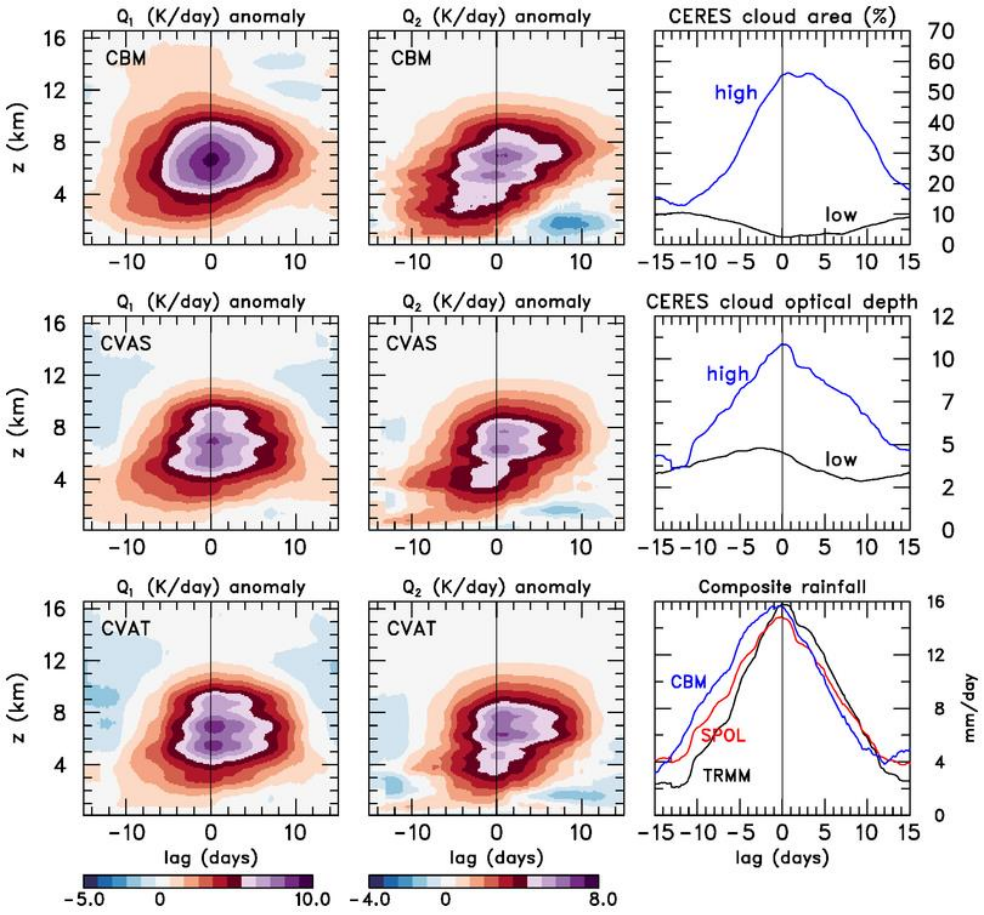


FIG. 13. Anomalous composite profiles as a function of lag in days relative to MJO peak SPOL rainfall for the period 2 October to 31 December: (left) $Q_1$, (middle) $Q_2$ for (top to bottom) CBM, CVAS, and CVAT analyses, respectively. Anomalies are computed by subtracting the composite mean at each vertical level. Scale for plots (K/day) is along bottom. The right-most column shows MJO composite lifecycles for (top to bottom) CERES cloud area and optical depth for high (< 440 hPa; blue) and low (> 680 hPa; black) clouds, and various rainfall estimates, respectively.

This article is protected by copyright. All rights reserved.

Accepted Article

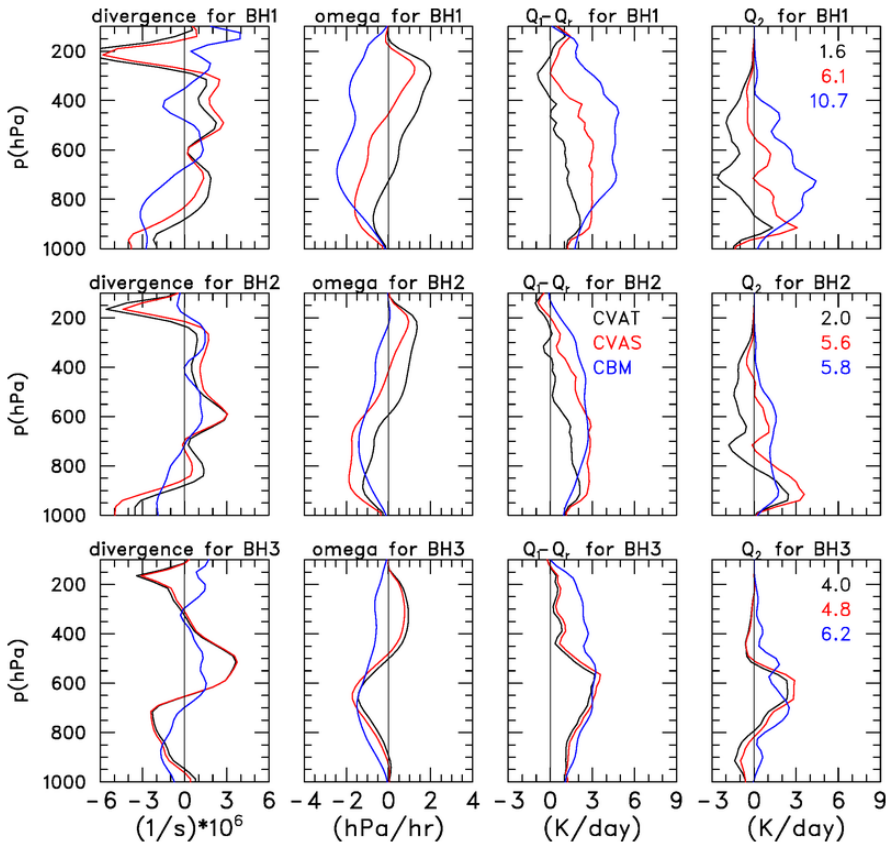


FIG. 14. Mean CVD vertical profiles from different analysis products (CVAT – black, CVAS – red, CBM – blue) for (left to right) divergence, vertical motion, apparent convective heating $Q_1$-$Q_R$, and apparent drying $Q_2$, respectively. Diagnostic profiles are shown for (top to bottom) the three BH MJO stages where BH1 goes from 10-12 October, BH2 goes from 8-16 November, and BH3 goes from 5-12 December, respectively. Period-mean rain rates (mm/day) are listed in right-most panels where CVAT is based on TRMM 3B42, CVAS is based on SPOL, and CBM is the $Q_2$ budget residual.

This article is protected by copyright. All rights reserved.

Accepted Article

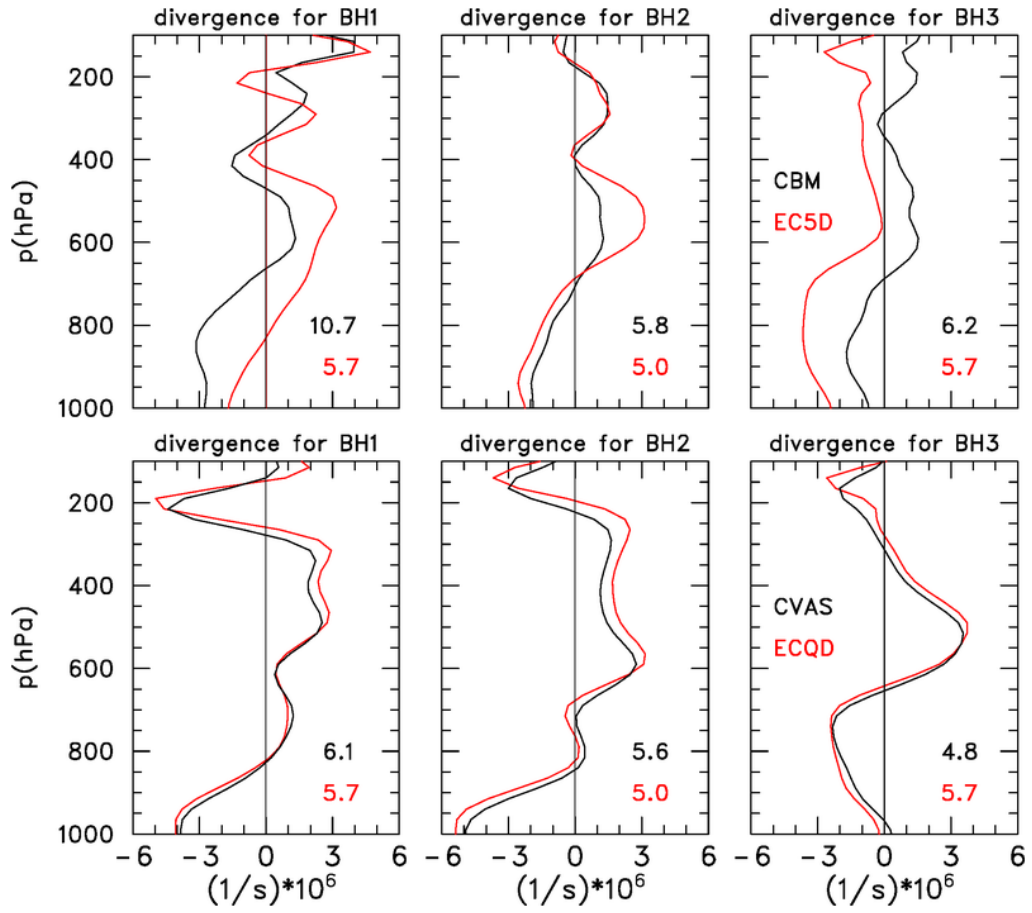


FIG. 15. Mean CVD vertical profiles of divergence from different analysis products for (left to right) three BH MJO periods. (top panels) CBM (black) and EC5D (red), (bottom panels) CVAS (black) and ECQD (red). Mean rain rates (mm/day) for each period and analysis are listed in bottom right corner of each panel.

This article is protected by copyright. All rights reserved.