

RESEARCH ARTICLE

Data assimilation errors due to incomplete accounting of balanced versus unbalanced moisture

H. Reed Ogrosky¹  | Alfredo N. Wetzel² | Leslie M. Smith^{3,4} | Samuel N. Stechmann^{3,4}

¹Department of Mathematics and Applied Mathematics, Virginia Commonwealth University, Richmond, Virginia, United States

²School of Mathematics and Statistics, The University of Sydney, Camperdown, New South Wales, Australia

³Department of Mathematics, University of Wisconsin-Madison, Madison, Wisconsin, United States

⁴Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, Wisconsin, United States

Correspondence

H. Reed Ogrosky, Department of Mathematics and Applied Mathematics, Virginia Commonwealth University, Richmond, VA 23284, United States
Email: hrogrosky@vcu.edu

Funding information

Simons Foundation, Grant/Award Number: 854116; US NSF DMS, Grant/Award Number: 1907667

Abstract

The notion of balance has played an important role in data assimilation. However, traditional ideas of balance have been based on dry dynamics, without moisture. Recently, the notion of balanced moisture has been formulated based on moist dynamics. Here, the goal is to investigate data assimilation strategies with moisture, and to assess the impact of accounting for moisture's balanced and unbalanced components. To illustrate the main ideas, examples are presented using a simple model consisting of linear, moist, shallow-water equations with stochastic forcing, and a Kalman filter in its stationary state for a data assimilation method akin to 3D-Var. If a traditional treatment of moisture is used, which does not properly decompose moisture into its balanced and unbalanced components, then the error in the recovered moisture is increased by up to 15% or more. These errors are most significant on smaller spatial scales, where inertia-gravity wave variance is larger, and in the presence of larger vertical gradients of the background moisture state, and in cases with larger observational noise. Underlying these results are some subtle theoretical aspects of the moist dynamical core and its dynamical operator, which is not symmetric; instead, it is self-adjoint with respect to a non-diagonal inner product, which is associated with conservation of moist energy. From the results here, and given the longstanding difficulties with data assimilation and forecasting of precipitation, it is possible that balanced versus unbalanced moisture may play a role in these difficulties, among many challenging aspects of water vapor, clouds, and precipitation.

KEYWORDS

3D-Var, data assimilation, geostrophic balance, humidity, inertia-gravity waves, Kalman filter, moisture, water vapor

1 | INTRODUCTION

The notion of balance has played an important role in data assimilation and weather forecasting. Even in the first numerical weather forecasts ever attempted (Richardson, 1922), the failure of the forecast has been in large part attributed to a failure to distinguish between balanced and unbalanced components (e.g., Kalnay, 2003; Platzman, 1967), associated with slow vortical/Rossby modes and fast inertia-gravity waves respectively. To improve data assimilation and forecasting, a great amount of attention has been paid to decompositions into balanced and unbalanced, or slow and fast, or vortical/Rossby and inertia-gravity wave components (e.g., Daley, 1993; Kalnay, 2003).

However, the traditional ideas of balance have been based on dry dynamics, without moisture (Bannister, 2021; Daley, 1993; Derber & Bouttier, 1999; Kalnay, 2003; Kasahara & Puri, 1981; Lorenc, 1981; Žagar *et al.*, 2015). Some ideas on balanced and unbalanced components of moisture have been introduced in the context of data assimilation, where one approach has been to introduce functional relationships or statistical regression relationships between a moisture variable (such as relative humidity or specific humidity) and a balanced variable (Bannister *et al.*, 2020; Berre, 2000; Chen *et al.*, 2013; Ingleby *et al.*, 2013; Michel *et al.*, 2011; Montmerle, 2012; Montmerle & Berre, 2010; Xiao & Sun, 2007). Although these relationships are based on sensible principles and physical intuition, they do not define a notion of balanced and unbalanced moisture that is based on different dynamical eigenmodes, such as the vortical/Rossby and inertia-gravity wave modes that often form the basis of balanced and unbalanced components of “dry” dynamics (velocity, temperature, etc.). To put dry and moist variables on the same footing, it would be desirable to define balance and imbalance from an explicit treatment of moist dynamics and its associated eigenmodes.

The notion of balanced moisture, based on moist dynamics, has recently been formulated by Smith and Stechmann (2017), and it is characterized by two (or more) quantities: not only potential vorticity (PV) but also a moist thermodynamic quantity called M . Wetzel *et al.* (2019, 2020) presented decompositions of moisture into its balanced and unbalanced components, which was done using an extension of PV inversion called PV-and- M inversion. Furthermore, the energy can also be decomposed in a way that draws out a component associated with balanced moisture (Kooloth *et al.*, 2023; Marsico *et al.*, 2019; Smith & Stechmann, 2017).

Here, the goal is to investigate data assimilation strategies with moisture and to assess the impact of accounting

for (or not accounting for) the balanced and unbalanced components of moisture.

The example system used here will be the linear, moist, shallow-water equations, in order to illustrate the main ideas. The aim with this system is to find analytical or semi-analytical solutions, so that errors can be assessed unambiguously. With this model system, we can specifically assess the errors introduced by an incomplete accounting of balanced versus unbalanced moisture, separate from other types of errors, such as model error or numerical error. Even in the linear setting, the moist dynamics presents several subtle properties, such as a non-symmetric dynamical operator and a non-diagonal inner product (see Sections 2.2 and 2.4); and it is valuable to first understand these subtleties in the linear setting. This model system also provides a moist analog of the dry shallow-water equations case that is helpful for illustrating the concepts of balanced and unbalanced components in a dry setting.

This study adds to the growing body of work on data assimilation with moisture, and in particular on simplified set-ups that can help to provide advances in understanding fundamental aspects. For instance, several idealized models have been introduced and studied with moisture and/or rainfall or precipitation (Chen & Majda, 2016; Goodliff *et al.*, 2017; Harlim & Majda, 2013; Janjić *et al.*, 2014; Janjić *et al.*, 2021; Kent *et al.*, 2017; Mou *et al.*, 2023; Robert & Künsch, 2017; Würsch & Craig, 2014; Žagar, 2012; Zaplotnik *et al.*, 2018; Zhu & Bannister, 2023). Here, we add to this body of work with an emphasis on balanced and unbalanced components of moisture. These types of simplified systems can provide helpful intuition and understanding for advancing more comprehensive systems for moist and/or convective-scale data assimilation (see, e.g., Dee & Da Silva, 2003; Gustafsson *et al.*, 2018; Bannister *et al.*, 2020, and other references listed above).

The remainder of the article is organized as follows. The model set-up is described in Section 2, including the decomposition of moist dynamics into its balanced and unbalanced components. Moist geostrophic adjustment is used as an illustrative example. The data assimilation set-up is described in Section 3. Results are shown in Section 4, and a concluding discussion is in Section 5.

2 | MODEL SET-UP: MOIST, LINEAR, SHALLOW-WATER EQUATIONS WITH DAMPING AND STOCHASTIC FORCING

In this section, the model equations are described in Section 2.1, and the balanced and unbalanced components of moisture are characterized in Sections 2.2 and 2.3.

An illustration is shown for the case of moist geostrophic adjustment. In particular, we highlight that two distinct eigenmodes contribute to balanced moisture: the vortical mode, which is widely known, and also a slow moist mode that we call the M mode and which has received less attention.

Furthermore, throughout Sections 2.2 and 2.3 and culminating in Section 2.4, we describe a subtlety that arises in the moist system and that has consequences for many aspects of data assimilation. Specifically, the dynamical operator is not symmetric, and the conserved energy is not the standard L^2 or Euclidean or mean-square energy. This has consequences for computing projections onto different eigenmodes, since it is no longer a standard orthogonal matrix that accomplishes the projection. Instead, a non-diagonal inner product appears; the dynamical operator is self-adjoint with respect to the non-diagonal inner product; and orthogonal projections can be computed if the non-diagonal inner product is utilized.

2.1 | Model equations

In prior studies, the dry, linear, shallow-water equations have been valuable as a simple system for analyzing data assimilation methodologies (e.g., Bube & Ghil, 1981; Kalnay, 2003; Talagrand, 1981). An advantageous feature of the linearized system is that its properties can often be understood with exact formulas and intuition.

Here, analogously, in order to understand fundamental aspects of data assimilation for a system that is moist, we use the moist, linear, shallow-water equations:

$$\frac{\partial u}{\partial t} - fv - \frac{gH_T}{\theta_0\pi} \frac{\partial}{\partial x} \theta = -D_u + \dot{z}_u, \quad (1a)$$

$$\frac{\partial v}{\partial t} + fu - \frac{gH_T}{\theta_0\pi} \frac{\partial}{\partial y} \theta = -D_v + \dot{z}_v, \quad (1b)$$

$$\frac{\partial \theta}{\partial t} - \frac{H_T}{\pi} \frac{d\theta_{bg}}{dz} \nabla_H \cdot \mathbf{u} = -D_\theta + \dot{z}_\theta, \quad (1c)$$

$$\frac{\partial q}{\partial t} - \frac{H_T}{\pi} \frac{dq_{bg}}{dz} \nabla_H \cdot \mathbf{u} = -D_q + \dot{z}_q. \quad (1d)$$

Equations 1a–1d arise from the three-dimensional primitive equations in describing the baroclinic vertical modes. In such derivations, the moisture variable can be interpreted as either a column-integrated water vapor mixing ratio (e.g., Khouider & Majda, 2006) or as midtropospheric moisture or midtroposphere-weighted vertical integral of moisture (Stechmann & Hottovy, 2017), and the form of the evolution equation for moisture is the same in either case. We also refer the reader to Stechmann *et al.* (2008) and Majda (2003) for details on their derivation.

The model variables are as follows: $\mathbf{u} = (u, v)^T$ is the horizontal velocity with Cartesian components u for the zonal and v for the meridional velocity, θ is the anomaly of potential temperature, and q is the anomaly of water vapor mixing ratio. The coordinate (x, y, z) is the three-dimensional position, with positive x pointing toward the east, positive y toward the north, and positive z in the upward direction. The symbol $\nabla_H = (\partial/\partial x, \partial/\partial y)$ denotes the horizontal gradient operator. In addition, f is the (constant) Coriolis parameter, g is acceleration due to gravity, H_T is a characteristic height for the tropopause, θ_0 is a reference potential temperature, $d\theta_{bg}/dz$ is a background vertical temperature gradient, and dq_{bg}/dz is a background vertical moisture gradient. Both $d\theta_{bg}/dz$ and dq_{bg}/dz are taken to be constants.

The dissipation terms D_u , D_v , D_θ , and D_q and stochastic forcing terms $\dot{z}_u$, $\dot{z}_v$, $\dot{z}_\theta$, and $\dot{z}_q$ are meant to represent a parametrization of nonlinear advection, unresolved turbulence, and other nonlinear processes, as is somewhat commonly done in idealized models, data-driven modeling, data assimilation, and prediction (e.g., Chen & Stechmann, 2019; DelSole, 2004; Harlim & Majda, 2008; Hottovy & Stechmann, 2015; Huang *et al.*, 2022). Each dissipation term D_u , D_v , D_θ , and D_q is a linear function of the state vector (u, v, θ, q) , and the stochastic forcing terms are not necessarily assumed to be independent here; both dissipation terms and stochastic forcing terms will be specified in terms of simpler diagonal modes in Equation (11) (see Section 2.2).

Note that Equations 1a–1d are in a midlatitude setting with constant Coriolis parameter $f > 0$. Also note that Equations 1a–1d include no precipitation, clouds, latent heating, or phase changes of water. It would be interesting to investigate other aspects of additional complexity in the future, such as tropical/equatorial waves, precipitation, cloud latent heating, and various nonlinearities, as many interesting challenges are known to arise due to these additional complexities (e.g., Zagar *et al.*, 2004a, 2004b; Zapolnik *et al.*, 2018; Zhu & Bannister, 2023). Here, we isolate the effect of balanced and unbalanced components of moisture, as an aspect that has received less attention in idealized studies of data assimilation, and which introduces interesting complexity even in a linear setting (e.g., see Sections 2.2 and 2.4).

It is also convenient to express Equations 1a–1d in Fourier space in matrix-vector form. Expressing $u(x, y, t) = \sum_k \sum_l \hat{u}(k, l, t) e^{i(kx+ly)}$, and similarly for v , θ , and q , the system of Equations 1a–1d may be recast as

$$\frac{\partial}{\partial t} \hat{\mathbf{u}} + \mathbf{A}_{kl} \hat{\mathbf{u}} = -\hat{\mathbf{D}} + \hat{\mathbf{Z}} \quad (2)$$

after projecting onto the wave number $\mathbf{k} = (k, l)$, where $\hat{\mathbf{u}} = (\hat{u}, \hat{v}, \hat{\theta}, \hat{q})^T$, $\hat{\mathbf{D}} = (\hat{D}_u, \hat{D}_v, \hat{D}_\theta, \hat{D}_q)^T$, and

$\hat{\mathbf{Z}} = (\hat{z}_u, \hat{z}_v, \hat{z}_\theta, \hat{z}_q)^T$ are the Fourier transform of the state vector, dissipation terms, and forcing terms respectively, and

$$\mathbf{A}_{kl} = \begin{pmatrix} 0 & -f & -ik\frac{gH_T}{\theta_0\pi} & 0 \\ f & 0 & -il\frac{gH_T}{\theta_0\pi} & 0 \\ -ik\tilde{\Theta} & -il\tilde{\Theta} & 0 & 0 \\ ik\tilde{Q} & il\tilde{Q} & 0 & 0 \end{pmatrix}. \quad (3)$$

The constant $\tilde{\Theta} = (H_T/\pi) d\theta_{bg}/dz$ is the vertical background temperature gradient scaled by the height of the troposphere. Therefore, $\tilde{\Theta}$ has units of temperature and is proportional to the total change in background temperature over the troposphere. Similarly, $\tilde{Q} = -(H_T/\pi) dq_{bg}/dz$ has units of moisture and is proportional to the total change in background water vapor over the troposphere.

2.2 | Eigenmodes, and balanced and unbalanced components

In this subsection, we define the balanced and unbalanced components of the moist system in Equations 1a–1d. The notion used here for balanced versus unbalanced is the characterization of the eigenvalues or oscillation frequencies as low frequency versus high frequency.

A self-contained description for the system in Equations 1a–1d is provided here for completeness. The concepts of balanced and unbalanced moisture here are similar to what was presented in earlier work (e.g., Smith & Stechmann, 2017; Wetzel *et al.*, 2019), although the details are different for every particular system. For settings that include additional realism beyond the simple model here (such as cloud latent heating, precipitation, three-dimensional fluid dynamics, anelastic equations, etc.), the reader is referred to the work of Smith and Stechmann (2017) and Wetzel *et al.* (2019, 2020) and the recent work by Remond-Tiedrez *et al.* (2024).

The system in Equation (2) may be diagonalized by means of an eigenvalue and eigenvector decomposition of matrix $\mathbf{A}_{kl}$. To accomplish this decomposition, it is convenient to exploit its symmetry properties and, in this way, obtain algebraically simpler expressions for the eigenvectors.

Though many physical systems (such as dry atmospheric dynamics) are manifestly in skew-symmetric form, note that the moist dynamics matrix $\mathbf{A}_{kl}$ in Equation (3) is neither skew-symmetric nor skew-Hermitian. Instead, it is skew-adjoint with respect to a non-diagonal inner product; see Section 2.4 for further discussion. For the present task of finding the eigenvalues and eigenvectors of $\mathbf{A}_{kl}$, the non-skew-symmetric property of $\mathbf{A}_{kl}$ suggests a two-step

approach: first, we use a similarity transformation to bring $\mathbf{A}_{kl}$ to real skew-symmetric form, and then we diagonalize.

Specifically, matrix $\mathbf{A}_{kl}$ given in Equation (3) is similar to a real skew-symmetric matrix under the transformation

$$\mathbf{A}_{kl} = (\mathbf{K}_{kl}\mathbf{P})^{-1}\mathbf{S}_\kappa(\mathbf{K}_{kl}\mathbf{P}), \quad (4)$$

where

$$\mathbf{S}_\kappa = \begin{pmatrix} 0 & -f & \kappa R_1 f & 0 \\ f & 0 & 0 & 0 \\ -\kappa R_1 f & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{K}_{kl} = \begin{pmatrix} \frac{ik}{\kappa} & \frac{il}{\kappa} & 0 & 0 \\ -\frac{il}{\kappa} & \frac{ik}{\kappa} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (5)$$

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{R_1 f}{\tilde{\Theta}} & 0 \\ 0 & 0 & \frac{\tilde{Q}}{\tilde{\Theta}c} & c \end{pmatrix}.$$

In these matrices, $\kappa = (k^2 + l^2)^{1/2}$ is the magnitude of the wave number, $R_1 = NH_T/\pi f = (1/f)(gH_T\tilde{\Theta}/\theta_0\pi)^{1/2}$ is the deformation radius for the first baroclinic mode, and $N^2 = (g/\theta_0) d\theta_{bg}/dz$ is the Brunt-Väisälä or buoyancy frequency. Notice that skew-symmetric matrix $\mathbf{S}_\kappa$ depends only on the magnitude of the wave number (k, l), hence the subscript κ . The unitary matrix $\mathbf{K}_{kl}$ depends on the direction of the wave number and represents a transformation from Cartesian velocities ($\hat{u}, \hat{v}$) to divergence and vorticity variables. Matrix $\mathbf{P}$ does not depend on the wave number whatsoever, and so $\mathbf{P}\hat{\mathbf{u}}$ simply recombines the $(\hat{\theta}, \hat{q})$ variables as a linear combination, $\hat{M}/c = (\tilde{Q}/\tilde{\Theta})\hat{\theta} + \hat{q}$, which will characterize a balanced component of moisture, as seen in more detail later herein. Notice that $\mathbf{P}$ contains the free constant c with units of length over time, which will be specified later.

Determining the eigenvalues and eigenvectors of the skew-symmetric matrix $\mathbf{S}_\kappa$ is significantly simpler than working with the original matrix $\mathbf{A}_{kl}$. So, we proceed by determining the spectral properties of $\mathbf{S}_\kappa$ and relate them to the original matrix using the transformation in Equation (4).

Since the two matrices are similar, the skew-symmetric matrix $\mathbf{S}_\kappa$ has the same eigenvalues as the original matrix $\mathbf{A}_{kl}$ (e.g., Horn & Johnson, 2012). The characteristic polynomial is

$$\lambda^2(\lambda^2 + \kappa^2 R_1^2 f^2 + f^2) = 0. \quad (6)$$

The eigenvalues are then $\lambda = 0, \pm if\omega_\kappa$, where $\omega_\kappa = (\kappa^2 R_1^2 + 1)^{1/2}$. In other words, the system has two sets of eigenvalues: the eigenvalue 0 with multiplicity 2 and the non-zero imaginary complex conjugate pair $\pm if\omega_\kappa$. The zero eigenvalues are associated with the balanced or

zero-frequency modes, whereas the imaginary eigenvalues are associated with the unbalanced or wave modes of the system.

The eigenvectors for the real skew-symmetric matrix $\mathbf{S}_\kappa$ are (for eigenvalues 0, 0, $-i\omega_\kappa$, $i\omega_\kappa$)

$$\mathbf{U}_\kappa = \begin{pmatrix} 0 & 0 & i\omega_\kappa & -i\omega_\kappa \\ 0 & \kappa R_1 & 1 & 1 \\ 0 & 1 & -\kappa R_1 & -\kappa R_1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\omega_\kappa} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}\omega_\kappa} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}\omega_\kappa} \end{pmatrix}, \quad (7)$$

where the diagonal matrix on the right acts to scale the eigenvectors into orthonormal form, and it is factored separately for visual clarity. The matrix of eigenvectors in Equation (7) is unitary ($\mathbf{U}_\kappa^{-1} = \mathbf{U}_\kappa^*$). Namely, the matrix $\mathbf{U}_\kappa$ satisfies the eigenvalue equation $\mathbf{S}_\kappa \mathbf{U}_\kappa = \mathbf{U}_\kappa \mathbf{\Lambda}_\kappa$ with $\mathbf{U}_\kappa \mathbf{U}_\kappa^* = \mathbf{U}_\kappa^* \mathbf{U}_\kappa = \mathbf{I}$, where $\mathbf{\Lambda}_\kappa = \text{diag}(0, 0, -i\omega_\kappa, i\omega_\kappa)$ and $\mathbf{I}$ is the identity matrix.

Therefore, by means of the similarity transformation, Equation (4), and eigenvector matrix, Equation (7), we have found the matrix transformation that takes a vector in physical space $\hat{\mathbf{u}} = (\hat{u}, \hat{v}, \hat{\theta}, \hat{q})$ into the diagonal modes of the eigenvector basis. In the eigenvector basis, the state vector is defined as a vector of complex amplitudes:

$$\hat{\boldsymbol{\phi}} = (\hat{\mathcal{M}}, \hat{\mathcal{P}\mathcal{V}}, \hat{\mathcal{W}}_1, \hat{\mathcal{W}}_2)^T. \quad (8)$$

The matrix for this transformation, $\hat{\boldsymbol{\phi}} = \mathbf{T}_{kl} \hat{\mathbf{u}}$, is

$$\mathbf{T}_{kl} = \mathbf{U}_\kappa^* \mathbf{K}_{kl} \mathbf{P}, \quad (9)$$

using $\mathbf{K}_{kl}$ and $\mathbf{P}$ from Equation (4) and $\mathbf{U}_\kappa$ from Equation (7). The formula $\mathbf{T}_{kl} = \mathbf{U}_\kappa^* \mathbf{K}_{kl} \mathbf{P}$ is a product of the $\mathbf{K}_{kl} \mathbf{P}$ matrix that transformed $\mathbf{A}_{kl}$ into skew-symmetric form and the $\mathbf{U}_\kappa^*$ matrix of eigenvectors of that skew-symmetric matrix. Notice that, by construction, the first two entries of the vector $\hat{\boldsymbol{\phi}} = (\hat{\mathcal{M}}, \hat{\mathcal{P}\mathcal{V}}, \hat{\mathcal{W}}_1, \hat{\mathcal{W}}_2)^T$ in Equation (8) represent the zero-frequency modes. We label the first entry $\hat{\mathcal{M}}$, or M-mode, as it is related to the physical moisture, and the second entry $\hat{\mathcal{P}\mathcal{V}}$, or PV-mode, as it is related to the PV. The last two entries of Equation (8) are the wave modes, so we label them $\hat{\mathcal{W}}_1$ and $\hat{\mathcal{W}}_2$.

The dynamical equations of motion can likewise be diagonalized. The matrix $\mathbf{T}_{kl}^{-1}$ is indeed an eigenvector basis for $\mathbf{A}_{kl}$; so it satisfies $\mathbf{T}_{kl} \mathbf{A}_{kl} \mathbf{T}_{kl}^{-1} = \mathbf{\Lambda}_\kappa$. In other words, using the transformation

$$\begin{pmatrix} \hat{\mathcal{M}} \\ \hat{\mathcal{P}\mathcal{V}} \\ \hat{\mathcal{W}}_1 \\ \hat{\mathcal{W}}_2 \end{pmatrix} = \mathbf{T}_{kl} \begin{pmatrix} \hat{u} \\ \hat{v} \\ \hat{\theta} \\ \hat{q} \end{pmatrix} \quad \text{or} \quad \hat{\boldsymbol{\phi}} = \mathbf{T}_{kl} \hat{\mathbf{u}}, \quad (10)$$

the differential equation in Equation (2) is then diagonalized to

$$\frac{\partial}{\partial t} \hat{\boldsymbol{\phi}} + \mathbf{\Lambda}_\kappa \hat{\boldsymbol{\phi}} = -\boldsymbol{\mu} \hat{\boldsymbol{\phi}} + \boldsymbol{\sigma} \hat{\mathbf{W}}, \quad (11)$$

where $\hat{\mathbf{W}} = (\hat{W}_\mathcal{M}, \hat{W}_{\mathcal{P}\mathcal{V}}, \hat{W}_{\mathcal{W}_1}, \hat{W}_{\mathcal{W}_2})^T$ are the components of the stochastic forcing for each diagonal mode and $\boldsymbol{\mu} = \text{diag}(\mu_\mathcal{M}, \mu_{\mathcal{P}\mathcal{V}}, \mu_{\mathcal{W}_1}, \mu_{\mathcal{W}_2})$ and $\boldsymbol{\sigma} = \text{diag}(\sigma_\mathcal{M}, \sigma_{\mathcal{P}\mathcal{V}}, \sigma_{\mathcal{W}_1}, \sigma_{\mathcal{W}_2})$ are diagonal matrices of dissipation and forcing coefficients. In Equation (11), $\boldsymbol{\mu} \hat{\boldsymbol{\phi}} = \mathbf{T}_{kl} \hat{\mathbf{D}}$ is the dissipation term and $\boldsymbol{\sigma} \hat{\mathbf{W}} = \mathbf{T}_{kl} \hat{\mathbf{Z}}$ is the stochastic forcing term.

To define the inverse transformation $\mathbf{T}_{kl}^{-1}$, first note that the matrix $\mathbf{A}_{kl}$ is not symmetric, and hence the inverse matrix of $\mathbf{T}_{kl}$ is not simply its transpose but is $\mathbf{T}_{kl}^{-1} = (\mathbf{P}^* \mathbf{P})^{-1} \mathbf{T}_{kl}^*$. As we shall discuss in Section 2.4, this is related to the fact that the natural inner product in physical space is weighted by the matrix block $\mathbf{P}^* \mathbf{P}$. For the moment, we proceed to find $\mathbf{T}_{kl}^{-1}$ without using the notion of this special weighted inner product space. As an alternative route, note that, explicitly, the transformation matrix $\mathbf{T}_{kl}$ in Equation (10) is

$$\begin{aligned} \mathbf{T}_{kl} &= \mathbf{U}_\kappa^* \mathbf{K}_{kl} \mathbf{P} \\ &= \begin{pmatrix} c & 0 & 0 & 0 \\ 0 & \frac{R_1}{\omega_\kappa} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}\kappa\omega_\kappa} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}\kappa\omega_\kappa} \end{pmatrix} \\ &\quad \times \begin{pmatrix} 0 & 0 & \frac{\bar{Q}}{\bar{\Theta}} & 1 \\ -il & ik & \frac{f}{\bar{\Theta}} & 0 \\ \omega_\kappa k - il & \omega_\kappa l + ik & -\frac{\kappa^2 R_1^2 f}{\bar{\Theta}} & 0 \\ -\omega_\kappa k - il & -\omega_\kappa l + ik & -\frac{\kappa^2 R_1^2 f}{\bar{\Theta}} & 0 \end{pmatrix}. \end{aligned} \quad (12)$$

The transformation matrix $\mathbf{T}_{kl}$ in Equation (12) has the advantage that it is written as the product of the two unitary matrices $\mathbf{U}_\kappa$, $\mathbf{K}_{kl}$ and the simple lower triangular, block matrix $\mathbf{P}$. Thus, the inverse transformation $\mathbf{T}_{kl}^{-1}$ is easily computed in terms of this product and gives

$$\begin{aligned} \mathbf{T}_{kl}^{-1} &= \mathbf{P}^{-1} \mathbf{K}_{kl}^* \mathbf{U}_\kappa \\ &= \begin{pmatrix} 0 & il & \omega_\kappa k + il & -\omega_\kappa k + il \\ 0 & -ik & \omega_\kappa l - ik & -\omega_\kappa l - ik \\ 0 & \frac{\bar{\Theta}}{R_1^2 f} & -\frac{\kappa^2 \bar{\Theta}}{f} & -\frac{\kappa^2 \bar{\Theta}}{f} \\ 1 & -\frac{\bar{Q}}{R_1^2 f} & \frac{\kappa^2 \bar{Q}}{f} & \frac{\kappa^2 \bar{Q}}{f} \end{pmatrix} \end{aligned}$$

$$\times \begin{pmatrix} \frac{1}{c} & 0 & 0 & 0 \\ 0 & \frac{R_1}{\omega_\kappa} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2\kappa\omega_\kappa}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2\kappa\omega_\kappa}} \end{pmatrix}. \quad (13)$$

To provide some physical intuition behind the mode variables $(\widehat{\mathcal{M}}, \widehat{\mathcal{P}\mathcal{V}}, \widehat{\mathcal{W}}_1, \widehat{\mathcal{W}}_2)^T$, we can express them in terms of vorticity and divergence as follows. Using the explicit transformation in Equation (12), we find that the normalized diagonalized variables in Equation (10) are

$$\begin{pmatrix} \widehat{\mathcal{M}} \\ \widehat{\mathcal{P}\mathcal{V}} \\ \widehat{\mathcal{W}}_1 \\ \widehat{\mathcal{W}}_2 \end{pmatrix} = \begin{pmatrix} c & 0 & 0 & 0 \\ 0 & \frac{R_1}{\omega_\kappa} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2\kappa\omega_\kappa}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2\kappa\omega_\kappa}} \end{pmatrix} \times \begin{pmatrix} \frac{\bar{Q}}{\bar{\Theta}}\hat{\theta} + \hat{q} \\ \hat{\zeta} + \frac{f}{\bar{\Theta}}\hat{\theta} \\ -i\omega_\kappa\hat{D} + \hat{\zeta} - \frac{\kappa^2 R_1^2 f}{\bar{\Theta}}\hat{\theta} \\ i\omega_\kappa\hat{D} + \hat{\zeta} - \frac{\kappa^2 R_1^2 f}{\bar{\Theta}}\hat{\theta} \end{pmatrix}, \quad (14)$$

where $\hat{\zeta} = ik\hat{v} - il\hat{u}$ is the relative vorticity and $\hat{D} = ik\hat{u} + il\hat{v}$ is the divergence. Since the free variable c has units of length over time, the normalized diagonalized variables have units of velocity.

We may equally define more physically motivated diagonalized variables simply by neglecting the normalization matrix in Equation (14). These modes are

$$\begin{pmatrix} \widehat{M} \\ \widehat{P\mathcal{V}} \\ \widehat{W}_1 \\ \widehat{W}_2 \end{pmatrix} = \begin{pmatrix} \frac{\bar{Q}}{\bar{\Theta}}\hat{\theta} + \hat{q} \\ \hat{\zeta} + \frac{f}{\bar{\Theta}}\hat{\theta} \\ -i\omega_\kappa\hat{D} + \hat{\zeta} - \frac{\kappa^2 R_1^2 f}{\bar{\Theta}}\hat{\theta} \\ i\omega_\kappa\hat{D} + \hat{\zeta} - \frac{\kappa^2 R_1^2 f}{\bar{\Theta}}\hat{\theta} \end{pmatrix}. \quad (15)$$

The modes $\widehat{P\mathcal{V}}$, $\widehat{W}_1$, and $\widehat{W}_2$ each have units of frequency, and $\widehat{M}$ has units of water vapor.

2.3 | Moisture and its balanced and unbalanced components

To obtain a projection onto balanced (zero-frequency) and unbalanced (wave) components for each variable, we can simply revert the transformation in Equation (10) and then read off the contributions of the balanced and unbalanced modes into each physical variable. Namely, balanced components are those associated with zero-frequency modes $\widehat{\mathcal{M}}$ and $\widehat{\mathcal{P}\mathcal{V}}$, whereas unbalanced components are those

associated with wave modes $\widehat{\mathcal{W}}_1$ and $\widehat{\mathcal{W}}_2$. In other words,

$$\begin{pmatrix} \hat{u}_b \\ \hat{v}_b \\ \hat{\theta}_b \\ \hat{q}_b \end{pmatrix} = \mathbf{T}_{kl}^{-1} \begin{pmatrix} \widehat{\mathcal{M}} \\ \widehat{\mathcal{P}\mathcal{V}} \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \hat{u}_u \\ \hat{v}_u \\ \hat{\theta}_u \\ \hat{q}_u \end{pmatrix} = \mathbf{T}_{kl}^{-1} \begin{pmatrix} 0 \\ 0 \\ \widehat{\mathcal{W}}_1 \\ \widehat{\mathcal{W}}_2 \end{pmatrix} \quad (16)$$

gives the balanced and unbalanced components of the physical variables respectively, where we use the subscript “b” to denote the balanced components and the subscript “u” to denote unbalanced components. Since the matrix $\mathbf{T}_{kl}^{-1}$ is an eigenvector basis for the matrix $\mathbf{A}_{kl}$, we may equally interpret the statements in Equation (16) as describing the balanced and unbalanced components in terms of the span of the zero-frequency and wave eigenvectors. For example, the balanced component of state vector $(\hat{u}, \hat{v}, \hat{\theta}, \hat{q})^T$ is given by the span of the zero-frequency eigenvectors (the first two vectors) of $\mathbf{T}_{kl}^{-1}$.

We may use Equation (13) to combine the equations in Equation (16) into the single equation

$$\begin{pmatrix} \hat{u}_b \\ \hat{v}_b \\ \hat{\theta}_b \\ \hat{q}_b \\ \hat{u}_u \\ \hat{v}_u \\ \hat{\theta}_u \\ \hat{q}_u \end{pmatrix} = \begin{pmatrix} 0 & ilR_1^2 & 0 & 0 \\ 0 & -ikR_1^2 & 0 & 0 \\ 0 & \frac{\bar{\Theta}}{f} & 0 & 0 \\ 1 & \frac{\bar{Q}}{f} & 0 & 0 \\ 0 & 0 & \frac{k\omega_\kappa + il}{\kappa^2} & \frac{-k\omega_\kappa + il}{\kappa^2} \\ 0 & 0 & \frac{l\omega_\kappa - ik}{\kappa^2} & \frac{-l\omega_\kappa - ik}{\kappa^2} \\ 0 & 0 & -\frac{\bar{\Theta}}{f} & -\frac{\bar{\Theta}}{f} \\ 0 & 0 & \frac{\bar{Q}}{f} & \frac{\bar{Q}}{f} \end{pmatrix} \times \begin{pmatrix} \frac{1}{c} & 0 & 0 & 0 \\ 0 & \frac{1}{\omega_\kappa R_1} & 0 & 0 \\ 0 & 0 & \frac{\kappa}{\sqrt{2\omega_\kappa}} & 0 \\ 0 & 0 & 0 & \frac{\kappa}{\sqrt{2\omega_\kappa}} \end{pmatrix} \begin{pmatrix} \widehat{\mathcal{M}} \\ \widehat{\mathcal{P}\mathcal{V}} \\ \widehat{\mathcal{W}}_1 \\ \widehat{\mathcal{W}}_2 \end{pmatrix}. \quad (17)$$

The equations above describe the transformation into eigenmode variables and the decomposition of each quantity into its balanced and unbalanced components. Note that cloud latent heating, phase changes, and precipitation can potentially have an effect on eigenmodes and non-linear interactions of eigenmodes (e.g., Remond-Tiedrez *et al.*, 2024; Wetzel *et al.*, 2019, 2020; Zhang *et al.*, 2021a, 2021b, 2022). However, balanced and unbalanced components of moisture arise even in the linear setting here, so it is valuable to first investigate balanced moisture in this simplest setting. Also, the system in Equations 1a–1d is the moist analog of the dry linear shallow-water equations,

which are a valuable setting for considering dry balance, and the moist version here provides a natural moist counterpart.

Of particular interest here are the balanced and unbalanced components of the moisture or water vapor $\hat{q}$. Equation (17) allows for the description of the balanced moisture in terms of normalized diagonal modes, unnormalized diagonal modes when using Equation (15), or simply in terms of physical variables when using Equation (14). Explicitly, in each case, the balanced moisture can be written as

$$\hat{q}_b = \frac{1}{c} \hat{\mathcal{M}} - \frac{1}{\omega_\kappa R_1} \frac{\tilde{Q}}{f} \hat{\mathcal{P}} \hat{\mathcal{V}} \quad (18a)$$

(in normalized diagonalized modes),

$$= \hat{M} - \frac{1}{\omega_\kappa^2} \frac{\tilde{Q}}{f} \hat{\mathcal{P}} \hat{\mathcal{V}} \quad (18b)$$

(in unnormalized diagonalized modes),

$$= \hat{q} - \frac{1}{\omega_\kappa^2} \frac{\tilde{Q}}{f} \left(\hat{\zeta} - \frac{\kappa^2 R_1^2 f}{\tilde{\Theta}} \hat{\theta} \right) \quad (18c)$$

(in physical variables).

Similarly, unbalanced moisture is given by

$$\hat{q}_u = \frac{\kappa}{\sqrt{2}\omega_\kappa} \frac{\tilde{Q}}{f} (\hat{\mathcal{W}}_1 + \hat{\mathcal{W}}_2) \quad (19a)$$

(in normalized diagonalized modes),

$$= \frac{1}{\omega_\kappa^2} \frac{\tilde{Q}}{f} \left(\hat{\zeta} - \frac{\kappa^2 R_1^2 f}{\tilde{\Theta}} \hat{\theta} \right) \quad (19b)$$

(in physical variables).

Lastly, we mention in passing that another fruitful approach to describe balanced and unbalanced components of variables arises from considering the inverse of Equation (16) or, equivalently, simply considering Equation (10). This approach has the benefit that we can obtain balanced and unbalanced components of the physical variables without direct reference to the transformation matrices, but rather only to the diagonalized modes, Equation (15), of the system once these are available. Inverting the equations in Equation (16) in terms of the modes, Equation (15), or, equivalently, projecting modes, Equation (15), onto the balanced and unbalanced components gives

$$\begin{pmatrix} \hat{M} \\ \hat{\mathcal{P}} \hat{\mathcal{V}} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\tilde{Q}}{\tilde{\Theta}} \hat{\theta}_b + \hat{q}_b \\ \hat{\zeta}_b + \frac{f}{\tilde{\Theta}} \hat{\theta}_b \\ -i\omega_\kappa \hat{D}_b + \hat{\zeta}_b - \frac{\kappa^2 R_1^2 f}{\tilde{\Theta}} \hat{\theta}_b \\ i\omega_\kappa \hat{D}_b + \hat{\zeta}_b - \frac{\kappa^2 R_1^2 f}{\tilde{\Theta}} \hat{\theta}_b \end{pmatrix} \quad \text{and}$$

$$\begin{pmatrix} 0 \\ 0 \\ \hat{\mathcal{W}}_1 \\ \hat{\mathcal{W}}_2 \end{pmatrix} = \begin{pmatrix} \frac{\tilde{Q}}{\tilde{\Theta}} \hat{\theta}_u + \hat{q}_u \\ \hat{\zeta}_u + \frac{f}{\tilde{\Theta}} \hat{\theta}_u \\ -i\omega_\kappa \hat{D}_u + \hat{\zeta}_u - \frac{\kappa^2 R_1^2 f}{\tilde{\Theta}} \hat{\theta}_u \\ i\omega_\kappa \hat{D}_u + \hat{\zeta}_u - \frac{\kappa^2 R_1^2 f}{\tilde{\Theta}} \hat{\theta}_u \end{pmatrix}. \quad (20)$$

The top equation of the bottom right matrix in Equation (20) gives a description of the unbalanced moisture in terms of the unbalanced potential temperature. More specifically, with the understanding that $\hat{\theta}_u = \hat{\theta} - \hat{\theta}_b$, $\hat{q}_u = \hat{q} - \hat{q}_b$, this equation can be interpreted as

$$\hat{q}_u = -\frac{\tilde{Q}}{\tilde{\Theta}} \hat{\theta}_u \quad \text{or} \quad \hat{q}_b = \hat{q} + \frac{\tilde{Q}}{\tilde{\Theta}} (\hat{\theta} - \hat{\theta}_b). \quad (21)$$

Equation (21) then says that the balanced component of the moisture is obtained by subtracting a suitable amount of the unbalanced potential temperature from the moisture.

To demonstrate these notions of dynamically balanced and unbalanced moisture, we now present an example of moist geostrophic adjustment. To do so, Equation (2) is solved numerically with no forcing or damping. An extended domain with $L = 12,000$ km was used in both x and y , with periodic boundary conditions. For an initial condition, a Gaussian bump was prescribed for θ , located in the center of the domain; all other variables were initialized to zero throughout the domain. The simulation was run for 15 hr. Figure 1 shows the initial condition for θ .

This setting is similar to previous studies of the atmospheric response to a heat source (e.g., Bretherton & Smolarkiewicz, 1989; Mapes, 1993; Stechmann & Majda, 2009), and in particular the study of Liu and Moncrieff (2004), which used the setting of an f -plane. The heat source is typically interpreted as cloud latent heating. The atmospheric response includes a transient inertio-gravity-wave response that propagates away from the source. Rather than applying a heat source in the model here, an initial warm bump was used.

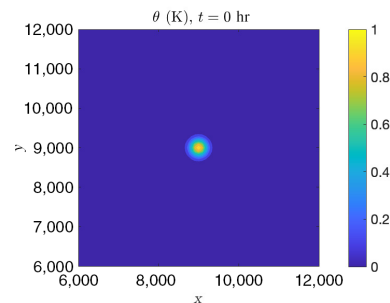


FIGURE 1 Initial condition for θ for illustration of moist geostrophic adjustment; u , v , and q were initialized to zero.

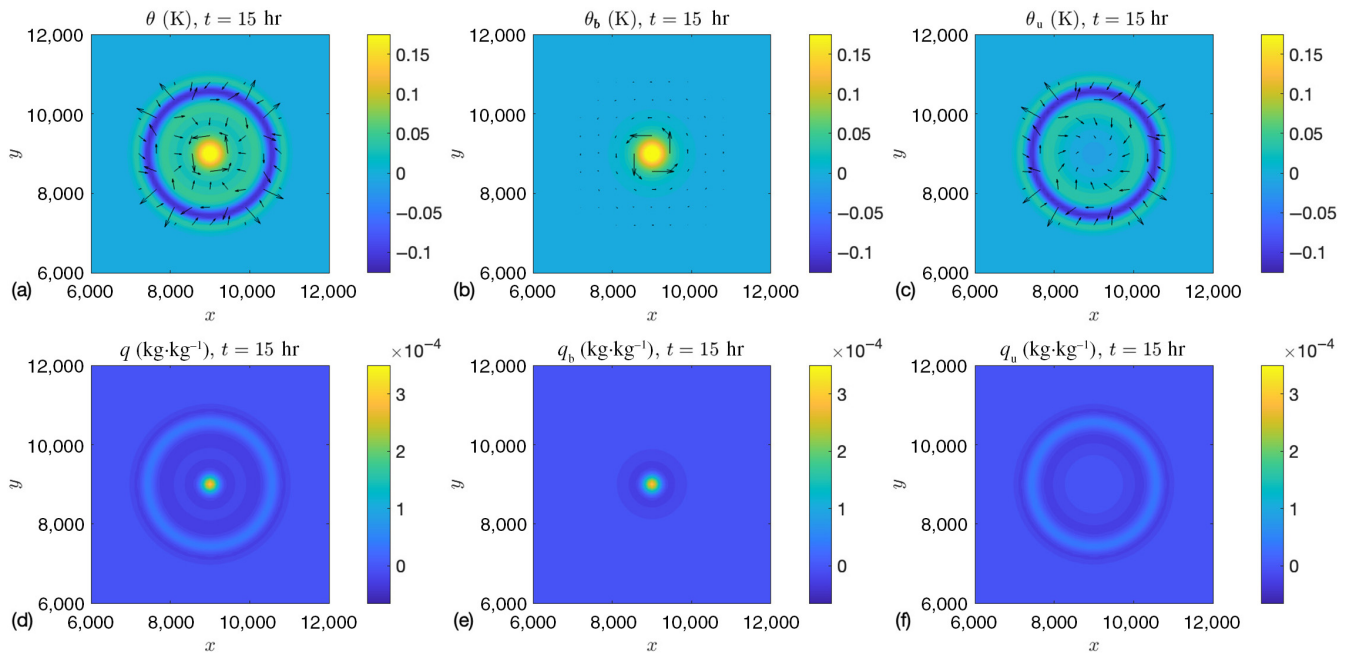


FIGURE 2 Illustration of balanced and unbalanced moisture. Solution to Equation (2) with no forcing or damping; initial condition shown in Figure 1. (a) Total, (b) balanced, and (c) unbalanced θ at $t = 15$ hr. (d) Total, (e) balanced, and (f) unbalanced q at $t = 15$ hr.

Figure 2a shows u , v , and θ at $t = 15$ hr. As part of the moist geostrophic adjustment process, a portion of the initial disturbance has propagated away from the domain center, and it appears as a ring shape in θ at $t = 15$ hr; at the same time, there also remains a steady portion of the bump at the domain center. Figure 2b,c shows the balanced and unbalanced components of u , v , and θ , obtained using Equation (16). The balanced component has isolated the steady portion of the bump at the domain center, whereas the unbalanced component captures the part of the initial disturbance that is propagating away from the domain center in the shape of a ring. Likewise, there is both a balanced and unbalanced component of moisture; Figure 2d–f shows the total, balanced, and unbalanced moisture.

As a brief comparison with observational data, we note that the balanced structure in Figure 2b,e is an idealization of many of the features seen in a composite midlatitude cyclone (Field & Wood, 2007). The basic structure in Figure 2b is a low in surface pressure, at the center of cyclonic winds. Note that a low in surface pressure is the same as a positive θ anomaly in the idealized model here, according to hydrostatic balance and the first vertical baroclinic mode structure associated with θ here. Also, balanced moisture is associated with the idealized cyclone in Figure 2e, similar to the positive anomalies of water vapor path, column relative humidity, and precipitation that are associated with a composite extratropical cyclone (Field & Wood, 2007). Of course, this is an

idealized comparison, but it helps to provide some context for thinking about the notion of balanced moisture.

What portion of the balanced moisture comes from $\widehat{\mathcal{M}}$ and what portion comes from $\widehat{\mathcal{P}\mathcal{V}}$? Figure 3a shows the balanced component of moisture; Figure 3b,c shows the portions q_M and q_{PV} that are attributable to the $\widehat{\mathcal{M}}$ and $\widehat{\mathcal{P}\mathcal{V}}$ modes respectively. In this particular case, q_M has a larger amplitude and narrower extent, whereas q_{PV} has a smaller amplitude and broader spatial extent. Figure 3d–f shows one-dimensional slices of the data in Figure 3a–c at $y = 9,000$. This figure serves to illustrate that balanced moisture has contributions from two different eigenmodes, including both the $\widehat{\mathcal{P}\mathcal{V}}$ mode and the $\widehat{\mathcal{M}}$ mode.

2.4 | Moist energy, inner product, and orthogonal projections

We now describe how the non-symmetric property of the moist dynamical operator has implications for the definition of the inner product, and therefore also implications for orthogonal projections. For instance, the standard Euclidean dot product must be replaced by a non-diagonal inner product.

These properties are related to the conserved energy, so we begin with considering the energy. Equation (11) for the diagonalized dynamical system suggests that the energy density $\mathcal{E}$ is in fact given by the standard (complex) Euclidean inner product $\mathcal{E} = \hat{\phi}^* \hat{\phi}$ of the diagonal

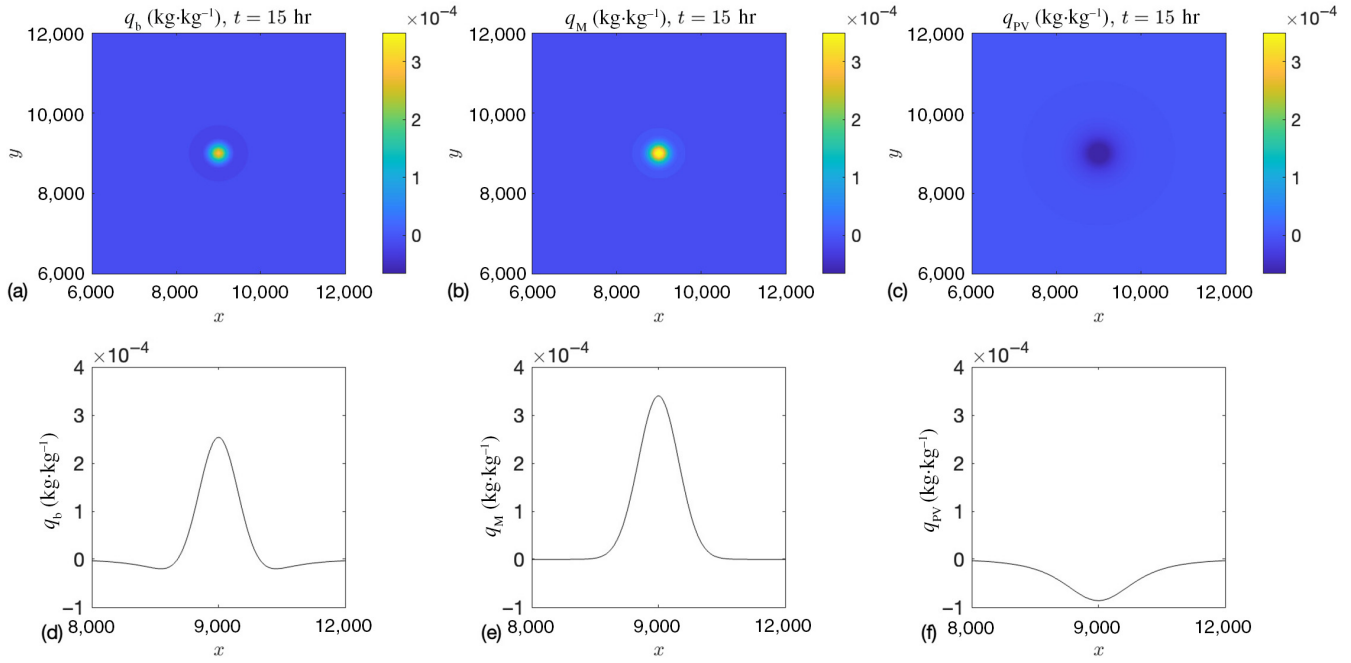


FIGURE 3 Balanced moisture has contributions from two eigenmodes. (a) Balanced moisture at $t = 15$ hr. (b) Balanced moisture attributable to $\widehat{\mathcal{M}}$ mode at $t = 15$ hr. (c) Balanced moisture attributable to $\widehat{\mathcal{P}}\widehat{\mathcal{V}}$ mode at $t = 15$ hr. (d)–(f) One-dimensional slices at $y = 9,000$ of (a)–(c) respectively.

modes $\widehat{\boldsymbol{\phi}} = (\widehat{\mathcal{M}}, \widehat{\mathcal{P}}\widehat{\mathcal{V}}, \widehat{\mathcal{W}}_1, \widehat{\mathcal{W}}_2)^T$. Namely, $\mathcal{E} = \widehat{\boldsymbol{\phi}}^* \widehat{\boldsymbol{\phi}}$ is conserved ($\partial \mathcal{E} / \partial t = 0$) in the absence of forcing and dissipation since $\mathbf{\Lambda}_\kappa$ is purely imaginary. The energy density of the system is then explicitly

$$\mathcal{E} = \widehat{\boldsymbol{\phi}}^* \widehat{\boldsymbol{\phi}} = |\widehat{\mathcal{M}}|^2 + |\widehat{\mathcal{P}}\widehat{\mathcal{V}}|^2 + |\widehat{\mathcal{W}}_1|^2 + |\widehat{\mathcal{W}}_2|^2 \quad (22)$$

(in normalized diagonalized modes),

$$= c^2 |\widehat{\mathcal{M}}|^2 + \frac{R_1^2}{\omega_\kappa^2} |\widehat{\mathcal{P}}\widehat{\mathcal{V}}|^2 + \frac{1}{2\kappa^2 \omega_\kappa^2} (|\widehat{\mathcal{W}}_1|^2 + |\widehat{\mathcal{W}}_2|^2) \quad (23)$$

(in unnormalized diagonalized modes).

Similarly, we may write the energy density grouped into diagonal modes as given in Equation (24) in terms of physical variables using the transformation in Equation (15) to obtain

$$\begin{aligned} \mathcal{E} = & c^2 \left| \frac{\tilde{Q}}{\tilde{\Theta}} \hat{\theta} + \hat{q} \right|^2 + \frac{R_1^2}{\omega_\kappa^2} \left| \hat{\zeta} + \frac{f}{\tilde{\Theta}} \hat{\theta} \right|^2 \\ & + \frac{1}{2\kappa^2 \omega_\kappa^2} \left| -i\omega_\kappa \hat{D} + \hat{\zeta} - \frac{\kappa^2 R_1^2 f}{\tilde{\Theta}} \hat{\theta} \right|^2 \\ & + \frac{1}{2\kappa^2 \omega_\kappa^2} \left| i\omega_\kappa \hat{D} + \hat{\zeta} - \frac{\kappa^2 R_1^2 f}{\tilde{\Theta}} \hat{\theta} \right|^2. \end{aligned} \quad (24)$$

Generally, the energy density $\mathcal{E}$ may be written in any coordinate system via the appropriate coordinate transformation. The transformation Equation (10) from physical space $\hat{\mathbf{u}} = (\hat{u}, \hat{v}, \hat{\theta}, \hat{q})^T$ to normalized diagonal modes

$\widehat{\boldsymbol{\phi}} = (\widehat{\mathcal{M}}, \widehat{\mathcal{P}}\widehat{\mathcal{V}}, \widehat{\mathcal{W}}_1, \widehat{\mathcal{W}}_2)^T$ gives rise to the energy density formula

$$\mathcal{E} = \widehat{\boldsymbol{\phi}}^* \widehat{\boldsymbol{\phi}} = (\mathbf{T}_{kl} \hat{\mathbf{u}})^* \mathbf{T}_{kl} \hat{\mathbf{u}} = \hat{\mathbf{u}}^* \mathbf{P}^* \mathbf{P} \hat{\mathbf{u}}, \quad (25)$$

where

$$\mathbf{P}^* \mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{R_1 f}{\tilde{\Theta}} & 0 \\ 0 & 0 & \frac{\tilde{Q}}{\tilde{\Theta}} c & c \end{pmatrix}^* \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{R_1 f}{\tilde{\Theta}} & 0 \\ 0 & 0 & \frac{\tilde{Q}}{\tilde{\Theta}} c & c \end{pmatrix}, \quad (26)$$

since $\mathbf{T}_{kl} = \mathbf{U}_\kappa^* \mathbf{K}_{kl} \mathbf{P}$ and $\mathbf{U}_\kappa, \mathbf{K}_{kl}$ are unitary. Explicitly, the energy density in physical space found in Equation (25) is then

$$\mathcal{E} = |\hat{u}|^2 + |\hat{v}|^2 + \frac{R_1^2 f^2}{\tilde{\Theta}^2} |\hat{\theta}|^2 + c^2 \left| \frac{\tilde{Q}}{\tilde{\Theta}} \hat{\theta} + \hat{q} \right|^2 \quad (27)$$

or

$$\begin{aligned} \mathcal{E} = & |\hat{u}|^2 + |\hat{v}|^2 + \left(\frac{R_1^2 f^2}{\tilde{\Theta}^2} + c^2 \frac{\tilde{Q}^2}{\tilde{\Theta}^2} \right) |\hat{\theta}|^2 \\ & + c^2 \frac{\tilde{Q}}{\tilde{\Theta}} (\hat{\theta}^* \hat{q} + \hat{q}^* \hat{\theta}) + c^2 |\hat{q}|^2. \end{aligned} \quad (28)$$

The formulas in Equation (25) suggest that we may consider the energy density simply as the magnitude of the

state vector in any given coordinate system with respect to a relevant inner product to measure its length. Specifically, the energy density is the magnitude of the normalized diagonal modes $(\hat{\mathcal{M}}, \hat{\mathcal{P}}\hat{\mathcal{V}}, \hat{\mathcal{W}}_1, \hat{\mathcal{W}}_2)^T$ using the standard Euclidean inner product. In physical space, however, the energy density is computed as the magnitude of the state vector $(\hat{u}, \hat{v}, \hat{\theta}, \hat{q})^T$ using a non-diagonal inner product with weight $\mathbf{P}^*\mathbf{P}$.

Another example of a non-diagonal inner product was described by Stechmann and Majda (2015) for a model of tropical intraseasonal oscillations, where again it is moisture that is the cause of the non-diagonal property of the inner product. Here, for the non-diagonal inner product in Equation (25), it is the weight matrix $\mathbf{P}^*\mathbf{P}$ that plays a key role.

More generally, therefore, for any two vectors $\mathbf{x}, \mathbf{y}$ in physical space, the appropriate inner product is $\mathbf{y}^*\mathbf{P}^*\mathbf{P}\mathbf{x}$. This fact becomes particularly salient when considering the eigenvectors of the original matrix $\mathbf{A}_{kl}$. The eigenvector basis $\mathbf{T}_{kl}^{-1}$ of the matrix $\mathbf{A}_{kl}$ consists of vectors that are not orthogonal using the Euclidean inner product. They do, however, satisfy the relation $(\mathbf{T}_{kl}^{-1})^*\mathbf{P}^*\mathbf{P}\mathbf{T}_{kl}^{-1} = \mathbf{I}$; orthogonality with respect to the inner product with weight $\mathbf{P}^*\mathbf{P}$. Notice that this also allows us to compute quickly the inverse of the matrix $\mathbf{T}_{kl}^{-1}$ of eigenvectors to be $\mathbf{T}_{kl} = (\mathbf{T}_{kl}^{-1})^*\mathbf{P}^*\mathbf{P}$.

We may also consider how to obtain the balanced and unbalanced components of the physical variables from the perspective of orthogonality with respect to the weighted inner product. Recall that the balanced and unbalanced components of $(\hat{u}, \hat{v}, \hat{\theta}, \hat{q})^T$ satisfy the formulas in Equation (20). Since $\mathbf{T}_{kl} = (\mathbf{T}_{kl}^{-1})^*\mathbf{P}^*\mathbf{P}$, Equation (20) together with Equation (10) suggest that balanced and unbalanced components of the physical variables are obtained simply via a projection onto the eigenvector basis $\mathbf{T}_{kl}^{-1}$. In particular, with respect to the weighted inner product, the balanced components of $(\hat{u}, \hat{v}, \hat{\theta}, \hat{q})^T$ are orthogonal to the wave eigenvectors, whereas unbalanced components are orthogonal to the zero-frequency eigenvectors. Notice, for example, that the bottom two rows of the top equation in Equation (20) are zero (balanced components are orthogonal to the wave eigenvectors), whereas the top two rows of the bottom equation in Equation (20) are zero (unbalanced components are orthogonal to the zero-frequency eigenvectors).

Lastly, the total energy is related to the covariance matrix of the state vector in the following manner. The total energy is given by the expected value of the energy density; that is, $\mathbb{E}[\mathcal{E}]$. Using the definition of the trace operator, for diagonal modes with mean zero, the total energy is

$$\mathbb{E}[\mathcal{E}] = \mathbb{E}[\hat{\boldsymbol{\phi}}^*\hat{\boldsymbol{\phi}}] = \text{tr}(\text{cov}(\hat{\boldsymbol{\phi}})), \quad (29)$$

since $\text{cov}(\hat{\boldsymbol{\phi}}) = \mathbb{E}[\hat{\boldsymbol{\phi}}\hat{\boldsymbol{\phi}}^*]$. Owing to the linearity of the expected value, the covariance in terms of diagonal modes is related to the covariance in physical space (e.g., Horn & Johnson, 2012) as

$$\text{cov}(\hat{\boldsymbol{\phi}}) = \mathbf{T}_{kl} \text{cov}(\hat{\mathbf{u}}) \mathbf{T}_{kl}^* \text{ or } \mathbf{T}_{kl}^{-1} \text{cov}(\hat{\boldsymbol{\phi}}) \mathbf{T}_{kl}^{-*} = \text{cov}(\hat{\mathbf{u}}). \quad (30)$$

Therefore, the total energy in terms of the covariance of the physical space variables is

$$\mathbb{E}[\mathcal{E}] = \text{tr}(\mathbf{T}_{kl} \text{cov}(\hat{\mathbf{u}}) \mathbf{T}_{kl}^*). \quad (31)$$

3 | DATA ASSIMILATION SET-UP

Now we examine the implications of balanced and unbalanced moisture, including the two balanced components (PV, M), for data assimilation. The recently introduced ideas of dynamically balanced and unbalanced moisture (Smith & Stechmann, 2017) have not been studied in any class of data assimilation algorithm. Therefore, though the role of dynamically balanced and unbalanced moisture could be explored in many data assimilation algorithms, it is sensible to first investigate them in a simple class of algorithms. Two such algorithms are the Kalman filter and three-dimensional variational (3D-Var) methods, which are discussed in Section 3.1. The benefit of the Kalman filter and 3D-Var is exact solutions, which give intuition and also unambiguous test results that are not corrupted by other sources of errors (numerical error, model error, etc.). Strategies for incorporating moisture in these methods are discussed in Section 3.2, and the set-up for the numerical experiments conducted here is outlined in Section 3.3.

The discrete-time version of the dynamical model in Equation (11) will be used in what follows. At discrete times $t_m = m\Delta t$, the discrete-time model is

$$\hat{\boldsymbol{\phi}}_{m+1} = \mathbf{F}\hat{\boldsymbol{\phi}}_m + \boldsymbol{\sigma}_{m+1}, \quad (32)$$

where the evolution operator $\mathbf{F}$ is diagonal and where linear damping is used for each mode with damping coefficients μ_M , μ_{PV} , and μ_{W_j} for $j \in \{1, 2\}$. The covariance of the stochastic forcing vector $\boldsymbol{\sigma}_{m+1}$ is denoted by $\mathbf{R} = \text{cov}(\boldsymbol{\sigma}_{m+1}) = \mathbb{E}[\boldsymbol{\sigma}_{m+1}\boldsymbol{\sigma}_{m+1}^*]$, and the mean is zero: $\mathbb{E}[\boldsymbol{\sigma}_{m+1}] = 0$. The stochastic forcing is white in time, so that $\mathbb{E}[\boldsymbol{\sigma}_m\boldsymbol{\sigma}_{m'}^*] = \delta_{mm'}\mathbf{R}$, where δ_{ij} is the Kronecker delta. The discrete-time version of the model in Equation (32) is equivalent to the continuous-time version of the model in Equation (11) and is obtained by analytically integrating Equation (11) forward in time by Δt time units. The notation throughout this section is consistent with that used in, for example, Majda and Harlim (2012)—with some

departures to reflect more traditional data assimilation notation as in Ide *et al.* (1997).

3.1 | Set-up for Kalman filter and 3D-Var as its stationary state

An observation $\hat{\psi}_{m+1}$ is made at each time $t_{m+1} = (m + 1)\Delta t$ and is described as

$$\hat{\psi}_{m+1} = \mathbf{H}\hat{\phi}_{m+1} + \sigma_{m+1}^o, \quad (33)$$

where $\mathbf{H}$ is an observation operator that will be specified later. Since the observation is linear and Gaussian, it also can be represented equivalently in terms of its mean and the covariance of the noise σ_{m+1}^o :

$$\bar{\hat{\psi}}_{m+1} = \mathbf{H}\bar{\hat{\phi}}_{m+1}, \quad (34a)$$

$$\mathbf{R}_{m+1}^o = \text{cov}(\sigma_{m+1}^o). \quad (34b)$$

For the filter, the state $\hat{\phi}_{m+1|m}$ of the system prior to any observations is given by

$$\hat{\phi}_{m+1|m} = \mathbf{F}\hat{\phi}_{m|m} + \sigma_{m+1}, \quad (35)$$

where $\hat{\phi}_{m|m}$ is the posterior state in which observations at $t = m\Delta t$ have been incorporated. Since the system is linear and Gaussian, the prior can be represented equivalently in terms of the mean and covariance:

$$\bar{\hat{\phi}}_{m+1|m} = \mathbf{F}\bar{\hat{\phi}}_{m|m}, \quad (36a)$$

$$\mathbf{R}_{m+1|m} = \mathbf{F}\mathbf{R}_{m|m}\mathbf{F}^* + \mathbf{R}. \quad (36b)$$

For this linear Gaussian system, the Kalman filter is an optimal filter (e.g., Anderson & Moore, 1979; Majda & Harlim, 2012), with the posterior mean and posterior covariance given by

$$\bar{\hat{\phi}}_{m+1|m+1} = \bar{\hat{\phi}}_{m+1|m} + \mathbf{K}_{m+1}(\hat{\psi}_{m+1} - \mathbf{H}\bar{\hat{\phi}}_{m+1|m}), \quad (37a)$$

$$\mathbf{R}_{m+1|m+1} = (\mathbf{I} - \mathbf{K}_{m+1}\mathbf{H})\mathbf{R}_{m+1|m}, \quad (37b)$$

where the Kalman gain matrix is

$$\begin{aligned} \mathbf{K}_{m+1} &= (\mathbf{R}_{m+1|m}^{-1} + \mathbf{H}^*(\mathbf{R}^o)^{-1}\mathbf{H})^{-1}\mathbf{H}^*(\mathbf{R}^o)^{-1} \\ &= \mathbf{R}_{m+1|m}\mathbf{H}^*(\mathbf{H}\mathbf{R}_{m+1|m}\mathbf{H}^* + \mathbf{R}^o)^{-1}. \end{aligned} \quad (38)$$

The long-time limit will also be considered here, in which case the initial conditions are no longer influential. In the long-time limit $m \rightarrow \infty$, under suitable stability conditions, the prior covariance, posterior covariance, and

Kalman gain will each converge toward constant matrices $\mathbf{R}_{\infty+1|\infty}$, $\mathbf{R}_{\infty|\infty}$, and $\mathbf{K}_{\infty}$ respectively:

$$\begin{aligned} \lim_{m \rightarrow \infty} \mathbf{R}_{m+1|m} &= \mathbf{R}_{\infty+1|\infty}, & \lim_{m \rightarrow \infty} \mathbf{R}_{m+1|m+1} &= \mathbf{R}_{\infty|\infty}, \\ \lim_{m \rightarrow \infty} \mathbf{K}_{m+1} &= \mathbf{K}_{\infty}. \end{aligned} \quad (39)$$

In the special case where $\mathbf{H}$, $\mathbf{R}$, and $\mathbf{R}^o$ are all diagonal, each of these limits is diagonal as well, with each diagonal entry corresponding to one scalar filter problem decoupled from the rest. In this case, each of the long-time prior $r_{\infty+1|\infty}$ and posterior $r_{\infty|\infty}$ covariance entries may be found from the positive root of

$$g^2 r_{\infty+1|\infty}^2 + (r^o - |F|^2 r^o - g^2 r) r_{\infty+1|\infty} - r^o r = 0, \quad (40a)$$

$$g^2 |F|^2 r_{\infty|\infty}^2 + (g^2 r + r^o - |F|^2 r^o) r_{\infty|\infty} - r^o r = 0, \quad (40b)$$

where g , r , r^o , and F are the corresponding diagonal entries of $\mathbf{H}$, $\mathbf{R}$, $\mathbf{R}^o$, and $\mathbf{F}$ respectively (Majda & Harlim, 2012).

A constant posterior covariance and constant Kalman gain are also sometimes used without the long-time-limit assumption, in the framework of a 3D-Var filter. (Note that the model here is not three-dimensional, but the term 3D-Var will still be used here, since the set-up here is meant to be an idealization of settings where 3D-Var is used.) The previous discussion indicates that 3D-Var is an optimal filter in the long-time limit (since the Kalman filter is an optimal filter, and the 3D-Var filter is the limit of the Kalman filter). As a result, if we restrict our attention to a long-time limit, then the 3D-Var filter can be used as a baseline for comparison against any simplified or suboptimal filters. Another benefit of 3D-Var is that the optimal 3D-Var algorithm is known for the model system here: it is the long-time ($m \rightarrow \infty$) asymptotic limit of the Kalman filter, in the stationary state of the stochastic dynamics (e.g., Majda & Harlim, 2012). This known (and analytically solvable) optimal filter will provide a base of comparison for other strategies.

In this stationary state, a 3D-VAR set-up is constructed by using the same equations in the Kalman filter in Equations (37a) and (38) except that the formula for the prior-to-posterior covariance update in Equation (37b) will be ignored. Instead, the prior covariance $\mathbf{R}_{m+1|m}$ is replaced by the background covariance $\mathbf{B}$. The analysis step is now given by

$$\begin{aligned} \bar{\hat{\phi}}_{m+1|m+1} &= \bar{\hat{\phi}}_{m+1|m} + \mathbf{B}\mathbf{H}^*(\mathbf{H}\mathbf{B}\mathbf{H}^* + \mathbf{R}^o)^{-1} \\ &\quad \times (\hat{\psi}_{m+1} - \mathbf{H}\bar{\hat{\phi}}_{m+1|m}). \end{aligned} \quad (41)$$

This 3D-Var filter will produce a (nearly) identical result to the Kalman filter, after a sufficiently long spin-up time

at the beginning of a simulation. One reason for proceeding from this point with 3D-Var filters instead of Kalman filters is because it allows a simpler set-up and a more straightforward comparison between the optimal, correct filter and suboptimal filters with suboptimality arising from erroneous prior covariance.

3.2 | Different treatments of moisture

In testing different data assimilation set-ups, we examine the implications of recognizing or not recognizing the proper decomposition using moist eigenvectors. In the tests that follow, two covariance matrices will be used: (a) the correct covariance matrix $\mathbf{B}$ and (b) an erroneous covariance matrix $\tilde{\mathbf{B}}$. The correct prior covariance matrix in Equation (41) is the long-time limit $\mathbf{B} = \mathbf{R}_{\infty+1|\infty}$ of the Kalman filter, Equation (39); note that in $\hat{\mathbf{u}}$ space, this prior covariance is

$$\text{cov}(\hat{\mathbf{u}}) = \mathbb{E}[\hat{\mathbf{u}}\hat{\mathbf{u}}^*] = \mathbf{T}_{kl}^{-1}\mathbf{B}(\mathbf{T}_{kl}^{-1})^*, \quad (42)$$

where $\mathbf{T}_{kl}$, defined in Equation (9), is the transformation matrix from $\hat{\mathbf{u}}$ space to $\hat{\boldsymbol{\phi}}$ space. We note that the variance in the $\hat{\mathcal{M}}$ mode contributes only to the $\hat{q}$ variance in $\hat{\mathbf{u}}$ space (whereas every other variable could affect every other variable).

An erroneous prior covariance will also be used, where an erroneous set of vectors is used in diagonalizing the prior covariance. A natural choice for this is $\tilde{\mathbf{T}}_{kl}^{-1}\mathbf{R}_{\infty+1|\infty}(\tilde{\mathbf{T}}_{kl}^{-1})^*$, where

$$\tilde{\mathbf{T}}_{kl} = \begin{pmatrix} c & 0 & 0 & 0 \\ 0 & \frac{R_1}{\omega_\kappa} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2\kappa\omega_\kappa}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2\kappa\omega_\kappa}} \end{pmatrix} \times \begin{pmatrix} 0 & 0 & 0 & 1 \\ -il & ik & \frac{f}{\Theta} & 0 \\ \omega_\kappa k - il & \omega_\kappa l + ik & -\frac{\kappa^2 R_1^2 f}{\Theta} & 0 \\ -\omega_\kappa k - il & -\omega_\kappa l + ik & -\frac{\kappa^2 R_1^2 f}{\Theta} & 0 \end{pmatrix}, \quad (43)$$

uses $\hat{q}$ instead of $\hat{\mathcal{M}}$ to describe the moisture eigenvector; that is, the only element different in $\mathbf{T}_{kl}$, Equation (12), and $\tilde{\mathbf{T}}_{kl}$, Equation (43), is the (1, 3) element. This covariance matrix corresponds to the assumption that “total” moisture is the second balanced mode and is uncorrelated with the dry variables. In the eigenvector basis space, this

erroneous prior covariance would then be

$$\begin{aligned} \tilde{\mathbf{B}} &= \mathbf{T}_{kl}\tilde{\mathbf{T}}_{kl}^{-1}\mathbf{R}_{\infty+1|\infty}(\tilde{\mathbf{T}}_{kl}^{-1})^*\mathbf{T}_{kl}^* \\ &= \begin{pmatrix} 1 & c\frac{1}{R_1}\frac{\tilde{Q}}{f\omega_\kappa} & -c\frac{\kappa}{\sqrt{2}}\frac{\tilde{Q}}{f\omega_\kappa} & -c\frac{\kappa}{\sqrt{2}}\frac{\tilde{Q}}{f\omega_\kappa} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &\quad \times \mathbf{R}_{\infty+1|\infty} \begin{pmatrix} 1 & 0 & 0 & 0 \\ c\frac{1}{R_1}\frac{\tilde{Q}}{f\omega_\kappa} & 1 & 0 & 0 \\ -c\frac{\kappa}{\sqrt{2}}\frac{\tilde{Q}}{f\omega_\kappa} & 0 & 1 & 0 \\ -c\frac{\kappa}{\sqrt{2}}\frac{\tilde{Q}}{f\omega_\kappa} & 0 & 0 & 1 \end{pmatrix}. \end{aligned} \quad (44)$$

From here onward, the free value c will be set to be $1 \text{ km} \cdot \text{s}^{-1}$ and will not appear explicitly.

We note that the entries of Equation (44) give some indication of the impact of using $\tilde{\mathbf{B}}$ instead of $\mathbf{B}$ in the limit of large or small spatial scales. In the long-wave limit $\kappa \rightarrow 0$ with other parameters fixed, $1/\omega_\kappa \rightarrow 1$ and so $\kappa/\omega_\kappa \rightarrow 0$, and the contribution from the unbalanced wave modes is expected to be negligible; thus, on large spatial scales (e.g., planetary and synoptic), any differences in the analysis produced using $\mathbf{B}$ versus $\tilde{\mathbf{B}}$ are attributable to the PV mode. On the other hand, in the limit $\kappa \rightarrow \infty$, we have $1/\omega_\kappa \rightarrow 0$ and $\kappa/\omega_\kappa \rightarrow 1/R_1$; thus, on these smaller spatial scales (e.g., mesoscale), any analysis differences will be due to the wave modes. The spatial scales in Table 1 were selected as they sit near the border of mesoscale and synoptic scales and thus lie at neither extreme; alternate cases exploring large or small spatial scales will be explored in Section 4.2.

3.3 | Set-up of numerical experiments

We now describe a set of numerical experiments using the model in Equation (11) rewritten here in component form:

$$d\hat{\mathcal{M}}(t) = -\mu_{\mathcal{M}}\hat{\mathcal{M}}(t) dt + \sigma_{\mathcal{M}} dW_{\mathcal{M}}(t), \quad (45a)$$

$$d\hat{\mathcal{P}}\hat{\mathcal{V}}(t) = -\mu_{\mathcal{P}\mathcal{V}}\hat{\mathcal{P}}\hat{\mathcal{V}}(t) dt + \sigma_{\mathcal{P}\mathcal{V}} dW_{\mathcal{P}\mathcal{V}}(t), \quad (45b)$$

$$d\hat{\mathcal{W}}_1(t) = (-\mu_{\mathcal{W}_1} - if\omega_\kappa)\hat{\mathcal{W}}_1(t) dt + \sigma_{\mathcal{W}_1} dW_{\mathcal{W}_1}(t), \quad (45c)$$

$$d\hat{\mathcal{W}}_2(t) = (-\mu_{\mathcal{W}_2} + if\omega_\kappa)\hat{\mathcal{W}}_2(t) dt + \sigma_{\mathcal{W}_2} dW_{\mathcal{W}_2}(t), \quad (45d)$$

where each of the $W_j(t)$, $j \in \{\mathcal{M}, \mathcal{P}\mathcal{V}, \mathcal{W}_1, \mathcal{W}_2\}$, is a Wiener process whose “derivative” is a Gaussian random

TABLE 1 Standard parameter values.

Parameter	Value (standard case)	Description
f	10^{-4} s^{-1}	Coriolis parameter at $\approx 45^\circ \text{ N}$
g	$0.0098 \text{ km} \cdot \text{s}^{-2}$	Acceleration due to gravity
H_T	10 km	Troposphere height
L	4,000 km	Simulation domain length (in both x and y)
$d\theta_{bg}/dz$	$3 \text{ K} \cdot \text{km}^{-1}$	Background potential temperature gradient
$-dq_{bg}/dz$	$0.0012 \text{ kg} \cdot \text{kg}^{-1} \cdot \text{km}^{-1}$	Background moisture gradient (negative)
θ_0	300 K	Reference potential temperature
(k, l)	(4, 4)	Zonal and meridional wave numbers
$\sigma_{\mathcal{M}}$	$5.06 \times 10^{-6} \text{ kg} \cdot \text{kg}^{-1} \cdot \text{s}^{-1/2}$	Stochastic forcing coefficient for $\widehat{\mathcal{M}}$
σ_{pV}	$4.29 \times 10^{-5} \text{ s}^{-3/2}$	Stochastic forcing coefficient for $\widehat{pV}$
$\sigma_{w_{1,2}}$	$3.04 \times 10^{-5} \text{ s}^{-3/2}$	Stochastic forcing coefficient for $\widehat{w}_{1,2}$
$\mu_{\mathcal{M}}$	$2.3 \times 10^{-5} \text{ s}^{-1}$	Damping coefficient for $\widehat{\mathcal{M}}$
μ_{pV}	$2.3 \times 10^{-5} \text{ s}^{-1}$	Damping coefficient for $\widehat{pV}$
$\mu_{w_{1,2}}$	$2.3 \times 10^{-5} \text{ s}^{-1}$	Damping coefficient for $\widehat{w}_{1,2}$
α	0.5	Observational noise parameter
γ	4	Additional noise factor for $\widehat{\mathcal{M}}$
Δt	10,800 s (3 hr)	Time between observations

variable with zero mean and unit variance. The exact solution of Equation (45a) is

$$\widehat{\mathcal{M}}(t) = e^{-\mu_{\mathcal{M}}t} \widehat{\mathcal{M}}(0) + \sigma_{\mathcal{M}} \int_0^t e^{-\mu_{\mathcal{M}}(t-s)} dW_{\mathcal{M}}(s), \quad (46)$$

where the mean and variance of $\widehat{\mathcal{M}}$ are

$$\begin{aligned} \overline{\widehat{\mathcal{M}}(t)} &= e^{-\mu_{\mathcal{M}}t} \widehat{\mathcal{M}}(0), \\ \text{Var}[\widehat{\mathcal{M}}(t)] &= \frac{\sigma_{\mathcal{M}}^2}{2\mu_{\mathcal{M}}} (1 - e^{-2\mu_{\mathcal{M}}t}), \end{aligned} \quad (47)$$

with similar solutions for the other three variables. If Δt is the time between observations, this corresponds to setting $\mathbf{F}$ and $\mathbf{R}$ in Equations (35) and (36a) and (36b) equal to

$$\mathbf{F} = \text{diag}(e^{-\mu_{\mathcal{M}}\Delta t}, e^{-\mu_{pV}\Delta t}, e^{(-\mu_{w_1} + i f \omega_k)\Delta t}, e^{(-\mu_{w_2} - i f \omega_k)\Delta t}), \quad (48a)$$

$$\mathbf{R} = \text{diag}(E_{\mathcal{M}}(1 - e^{-2\mu_{\mathcal{M}}\Delta t}), E_{pV}(1 - e^{-2\mu_{pV}\Delta t}), E_{w_1}(1 - e^{-2\mu_{w_1}\Delta t}), E_{w_2}(1 - e^{-2\mu_{w_2}\Delta t})), \quad (48b)$$

where $E_j = \sigma_j^2/2\mu_j$ is the climatological variance for $j \in \{\mathcal{M}, pV, w_1, w_2\}$.

In the numerical experiments, a single realization of Equations (45a)–(45d) was taken to be the system's true state. Table 1 shows the parameter values used. A $4,000 \times 4,000 \text{ km}^2$ domain was used, with wave number $k = l = 4$,

corresponding to a global zonal wave number of $k_x \approx 30$ in the midlatitudes. Identical damping time-scales of two days were used for all modes, with the stochastic forcing for the dry modes chosen to ensure that the dry energy was evenly split between $\widehat{pV}$ and the wave modes $\widehat{w}_1, \widehat{w}_2$ as seen in reanalysis data on these spatial scales: $E_{pV} \approx E_{w_1} + E_{w_2}$ (Žagar *et al.*, 2015). Stochastic forcing for moisture was then chosen to produce reasonable climatological variance in moisture.

Next, artificial observations were constructed by adding noise to the true values every $\Delta t = 3 \text{ hr}$. Experiments were conducted with two types of observations. In the first case (denoted EIG-OBS here), the observed quantities were the eigenvector variables $\widehat{\phi} = (\widehat{\mathcal{M}}, \widehat{pV}, \widehat{w}_1, \widehat{w}_2)^T$, and the observations of these four quantities were assumed to be statistically independent, so that the observation operator and observational noise covariance matrix (taken to be independent of time here) are diagonal:

$$\begin{aligned} \mathbf{H} &= \mathbf{I}, \\ \mathbf{R}_{m+1}^0 &= \mathbf{R}^0 = \text{diag}((\sigma_{\mathcal{M}}^0)^2, (\sigma_{pV}^0)^2, (\sigma_{w_1}^0)^2, (\sigma_{w_2}^0)^2), \end{aligned} \quad (49)$$

where $\mathbf{I}$ is the identity matrix. The diagonal entries of $\mathbf{R}^0$ were chosen according to

$$\begin{aligned} (\sigma_{pV}^0)^2 &= \alpha E_{pV}, & (\sigma_{w_{1,2}}^0)^2 &= \alpha E_{w_{1,2}}, \\ (\sigma_{\mathcal{M}}^0)^2 &= \alpha \gamma E_{\mathcal{M}}, \end{aligned} \quad (50)$$

where $\alpha \in \mathbb{R}^+$ is a parameter that sets the noise level, and $\gamma \geq 1$ is an additional parameter meant to reflect the possibly higher degree of uncertainty present in moisture observations. Because the observation matrix $\mathbf{H}$ and observational noise covariance $\mathbf{R}^o$ are both diagonal, this EIG-OBS case is a collection of independent, complex-valued “scalar” Kalman filters, one for each eigenmode. Owing to this decoupling, the analytical solution, Equations (39) and (40a) and (40b), the difference in RMSE, for the long-time ($m \rightarrow \infty$) asymptotic limit may be used.

In the second case (denoted here PHYS-OBS), the observed quantities were the physical variables $\hat{\mathbf{u}} = (\hat{u}, \hat{v}, \hat{\theta}, \hat{q})^T$ in Fourier space, and the observations of these four physical variables were assumed to be statistically independent, so that

$$\mathbf{H} = \mathbf{T}^{-1},$$

$$\mathbf{R}_{m+1}^o = \mathbf{R}_p^o = \text{diag}((\sigma_u^o)^2, (\sigma_v^o)^2, (\sigma_\theta^o)^2, (\sigma_q^o)^2). \quad (51)$$

Here, the subscript “P” denotes the PHYS-OBS case, and the values in $\mathbf{R}_p^o$ in Equation (51) were chosen—with Equation (30) in mind—to be the diagonal elements of the matrix $\mathbf{T}_{kl}^{-1} \mathbf{R}^o (\mathbf{T}_{kl}^{-1})^*$, with $\mathbf{R}^o$ defined in Equation (49). This PHYS-OBS case has the advantage of using independent observations of physical variables. However, the Kalman filter problems remain coupled in this PHYS-OBS case, and the long-time ($m \rightarrow \infty$) asymptotic limit does not have an analytical solution. As a result, some transparency is lost, and the inner workings of the methods may be less clear. For this reason, we elect to focus primarily on the EIG-OBS case here, where observations were made in eigenmode space; results for the PHYS-OBS case will be discussed in the [Supporting Information](#) and will be shown to be largely similar to those of the EIG-OBS case.

The Kalman filter was then used to assimilate these observations, as in Section 3.1, for a sufficiently long time (50 days) so that the system reached a statistical steady state, where the prior covariance $\mathbf{R}_{m+1|m}$ approximately reached its $m \rightarrow \infty$ asymptotic value $\mathbf{R}_{\infty+1|\infty}$. Next, two 3D-Var tests were run, one with the correct covariance matrix $\mathbf{B} = \mathbf{R}_{\infty+1|\infty}$ and one with the erroneous covariance matrix $\tilde{\mathbf{B}}$ in Equation (44). Each test consisted of an additional 50,000 assimilation cycles. The resulting posterior values were then compared with the true signal.

4 | RESULTS

In this section, the results of the tests described in Section 3.3 are presented in order to illustrate the impact of different strategies for incorporating moisture into 3D-Var. Specifically, the difference between the true state and the posterior states found by solving Equation (41) using (a)

$\mathbf{B} = \mathbf{R}_{\infty+1|\infty}$ and using (b) $\tilde{\mathbf{B}}$ from Equation (44) will be shown. Section 4.1 discusses the impact of varying model parameters in the numerical tests, and some analytical results valid in the asymptotic limit of large observational time are given. Increases in the error of 10% or more are seen if the erroneous covariance matrix $\tilde{\mathbf{B}}$ is used instead of the correct covariance $\mathbf{B}$. Section 4.2 explores the error increases (EIs) seen on small (or large) spatial scales at which the portion of energy present in balanced modes is small (or large) in reanalysis data.

4.1 | Equal dry balanced and unbalanced energy

We first consider the case where the dry energy is split equally between the unbalanced wave modes and balanced $\widehat{\mathcal{P}}\widehat{\mathcal{V}}$ as part of the standard set-up described in Section 3.3. Parameter values used are those in Table 1; this will be referred to as the standard case. A 10-day period of the numerical simulation is shown in Figure 4, with the true state of the eigenmode variables shown in black. Real parts are shown in Figure 4a–d; real and imaginary parts are shown in Figure 4e–h for the last 2.5 days of the simulation. Note that, with the standard parameter values used here, the wave period $2\pi/\omega_K \approx 6$ hr is roughly twice that of the time between observations. Noisy observations, taken every 3 hr, and the results of the analysis using $\mathbf{B}$ ($\tilde{\mathbf{B}}$) are shown in blue (red). The same simulation, shown in physical space variables $\hat{u}, \hat{v}, \hat{\theta}, \hat{q}$, is shown in the [Supporting Information](#).

At most times, noticeable differences in the analyses of $\widehat{\mathcal{M}}$ may be seen, whereas there is little discernible difference in the two analyses in $\widehat{\mathcal{W}}_1$, $\widehat{\mathcal{W}}_2$, and $\widehat{\mathcal{P}}\widehat{\mathcal{V}}$ in Figure 4. Note that only the real part of each mode is shown in Figure 4a–d; thus, there are some times when it may appear that the dry (or moist) analysis is closer to the truth than the moist (dry), but the distance in the complex plane may in fact be greater between the truth and the dry (moist) analysis. It is also important to note that there can be times when the dry analysis is closer to the truth. Supporting Information Figure S8 shows a 40-day period from another simulation. The moist analysis is closer to the truth than the dry analysis roughly two-thirds of the time.

Figure 5a–c shows the root-mean-square error (RMSE) for the simulation in Figure 4 (closed circles), as well as for a variety of values of vertical background moisture gradient dq_{bg}/dz ; the RMSE was calculated over the entire 3D-Var period for each of the two analyses. The EI due to the erroneous background error covariance matrix is seen most clearly in state estimates of $\widehat{\mathcal{M}}$, consistent with Figure 4a–d. The EI is also largest for large values of dq_{bg}/dz ; this may be expected ahead of time, since $\widehat{\mathcal{M}} \approx \hat{q}$

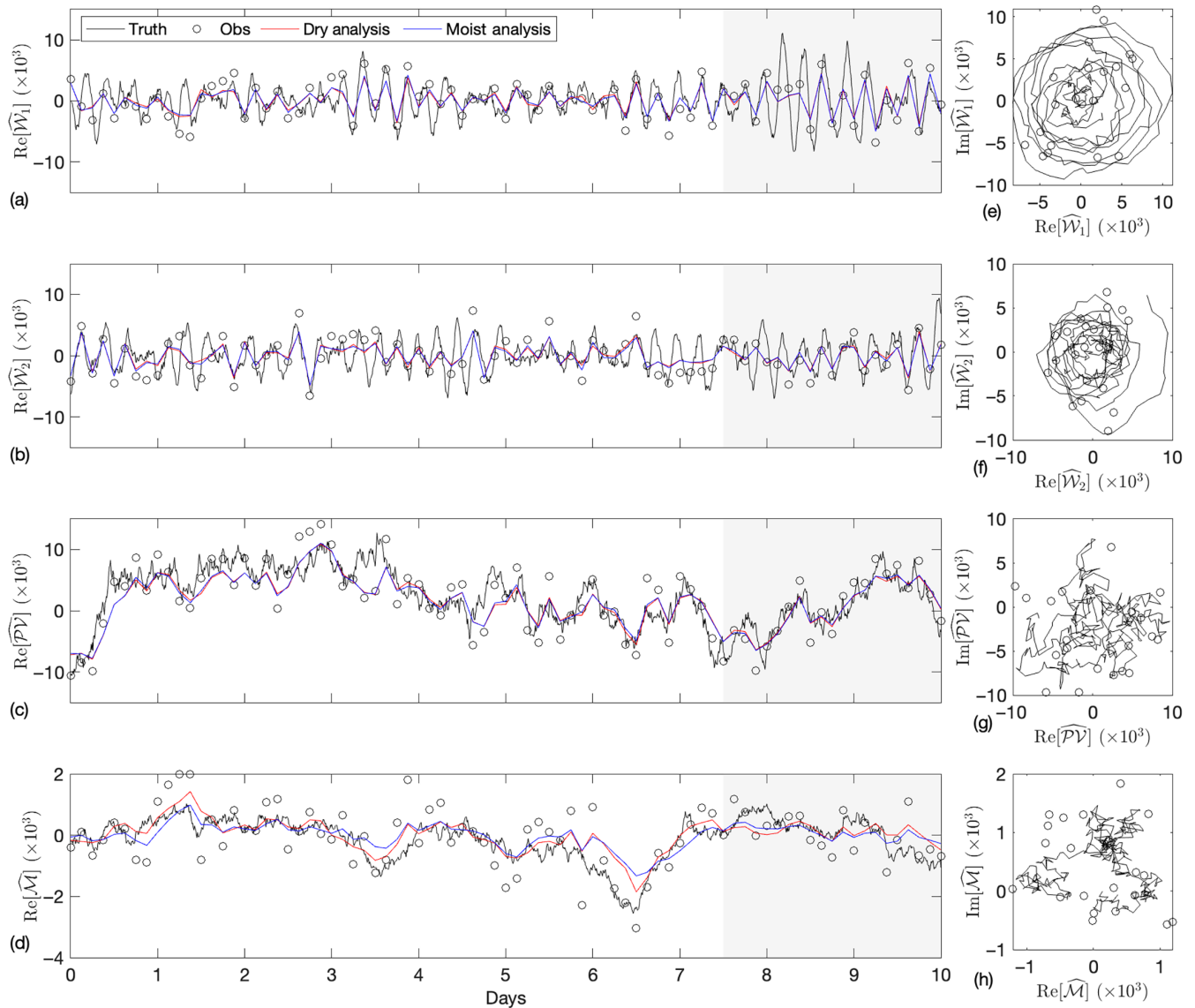


FIGURE 4 Final 10 days from 3D-Var with standard parameter values in Table 1. (a)–(d) Real part of the true signal (solid black), observations (black circles), moist analysis, Equation (41) (blue line) and dry analysis, Equation (41), except $\mathbf{B}$ replaced by $\tilde{\mathbf{B}}$ (red line) for eigenmode variables $\widehat{W}_1$, $\widehat{W}_2$, $\widehat{PV}$, and $\widehat{M}$. (e)–(h) Real and imaginary parts of true signal and observations for the last 2.5 days—shaded region of (a)–(d).

for dq_{bg}/dz near zero, and the background error covariance matrices $\mathbf{B}$ and $\tilde{\mathbf{B}}$ used in the two analyses are nearly identical.

Figure 5d shows the EI, defined here as

$$EI = \frac{\widetilde{\text{RMSE}} - \text{RMSE}}{\text{RMSE}}, \quad (52)$$

where $\text{RMSE}(\widetilde{\text{RMSE}})$ refers to the RMSE obtained by comparing the results found with $\mathbf{B}$ ($\tilde{\mathbf{B}}$) with the true signal. The EI in estimating $\widehat{M}$ exceeds 10% for the highest value of dq_{bg}/dz . Figure 5e–h shows the EI in state estimation of physical variables. There is up to 10% EI in estimation of $\hat{q}$, 3% increase in $\hat{\theta}$, and negligible increase in velocities.

What impact does observation frequency have on the error and the EI? To answer this, the previously mentioned tests were repeated for a variety of values of $\mu_M \Delta t$. Recall that μ_M is the damping rate of $\widehat{M}$, so $\mu_M \Delta t$ is the ratio of the observation time and decorrelation time of $\widehat{M}$. Values of $\mu_M \Delta t$ are denoted by the circles in Figure 6, which illustrates the damping factor $\exp(-\mu_M \Delta t)$ for different $\mu_M \Delta t$ values. These values span a large range in order to include cases where $\mu_M \Delta t$ is much smaller than the $\widehat{M}$ mode's decorrelation time, and cases where it is much larger. The standard case is shown by the shaded black circle.

Figure 7 shows the RMSE, the difference in RMSE, and the EI in $\widehat{M}$ for various values of $\mu_M \Delta t$ and three values of observational noise strength parameter α ; the standard

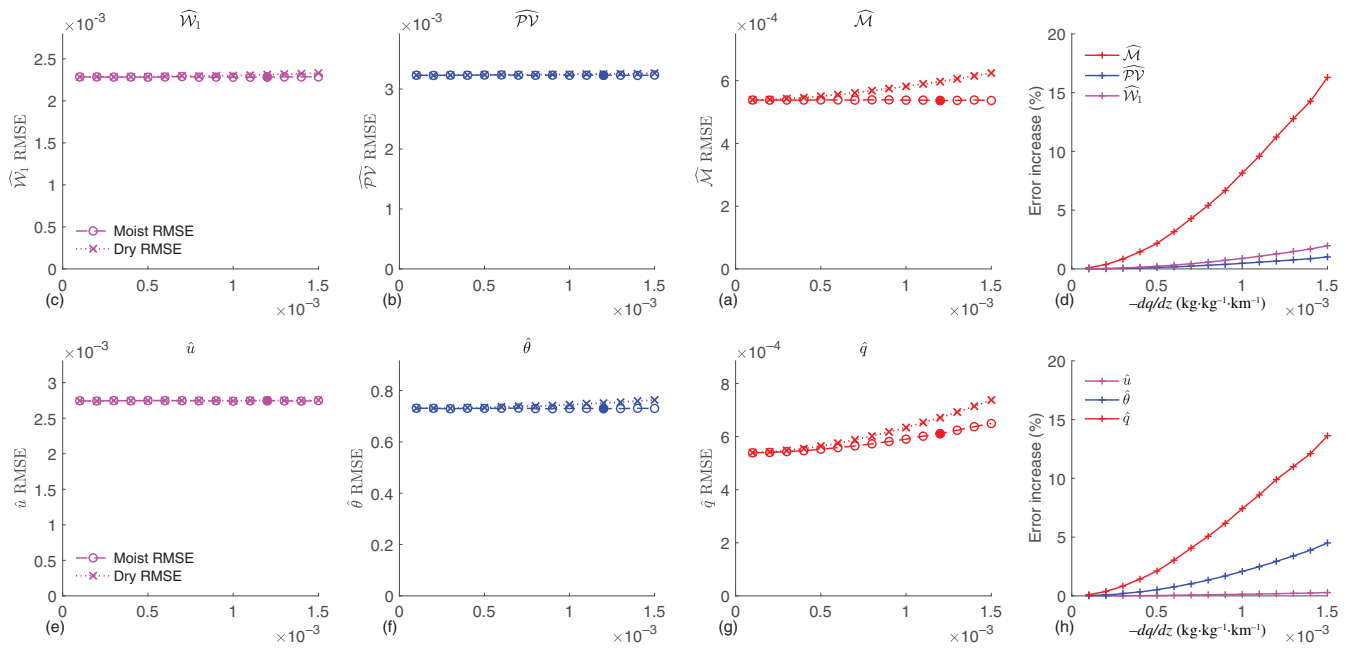


FIGURE 5 (a)–(c) Root-mean-square error (RMSE) for mode variables for various values of dq_{bg}/dz . $\widehat{W}_2$ results are nearly identical to those of $\widehat{W}_1$ and are omitted. (d) Error increase for mode variables due to using $\widehat{\mathbf{B}}$, Equation (44), instead of $\mathbf{B}$, Equation (41). (e)–(h) Same as (a)–(d) but for physical variables $\hat{u}$, $\hat{\theta}$, and $\hat{q}$; $\hat{v}$ results are nearly identical to those of $\hat{u}$ and are omitted. Error increase can be up to 10% or larger in moist eigenmode $\widehat{M}$ in (d) and in moisture $\hat{q}$ in (h).

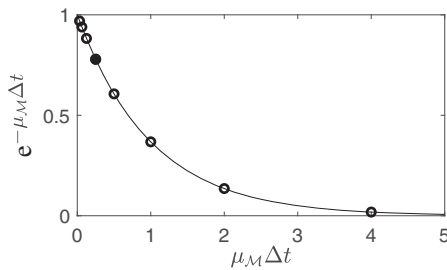


FIGURE 6 $e^{-\mu_M \Delta t}$ for parameters used in Figure 7. Circles correspond to values of $\mu_M \Delta t$ used in Figure 7.

case is shown with the shaded red circle. As the time between observations increases, and as the observational noise parameter α increases, the RMSE in both methods increases as expected. Interestingly, the EI decreases as observation time increases, with both the RMSE and EI each appearing to approach some value in the limit of large time between observations, shown by the gray lines. Note that the leftmost values in Figure 7a correspond to an observation time just greater than 5 min and the rightmost values in Figure 7a correspond to an observation time of 2 days (with very little change between 12 hr and 2 days). Supporting Information Figure S6 shows similar results but for $dq_{bg}/dz = -0.6 \text{ g} \cdot \text{kg}^{-1} \cdot \text{km}^{-1}$; as may be expected from Figure 5, the RMSE difference and EI are both smaller than seen in Figure 7.

The error in this large-observation-time limit may be found analytically by noting that for observation times much larger than the modes' decorrelation times, we then have that the model dynamical operator $\mathbf{F} \rightarrow \mathbf{0}$ (or $|\mathbf{F}| \rightarrow 0$ in the scalar case), and the only knowledge the model contributes to state estimation is the climatological variance. In this case, the prior mean $\bar{\phi}_{m+1|m} \approx 0$ can be neglected, so that Equation (41) can be written

$$\bar{\phi}_{m+1|m+1} \approx \mathbf{B}\mathbf{H}^*(\mathbf{H}\mathbf{B}\mathbf{H}^* + \mathbf{R}^0)^{-1}\hat{\psi}_{m+1}. \quad (53)$$

The mean-square error (MSE) may be found in the usual way. Some details of the calculation are shown in the Supporting Information, and a summary of the result is as follows. In the case of the correct covariance matrix $\mathbf{B}$, the MSE reduces to

$$\text{MSE} \approx [\mathbf{I} - \mathbf{B}\mathbf{H}^*(\mathbf{H}\mathbf{B}\mathbf{H}^* + \mathbf{R}^0)^{-1}\mathbf{H}]\mathbf{B}. \quad (54)$$

Thus, in this case with $\mathbf{H} = \mathbf{I}$ and with diagonal $\mathbf{R}$ and $\mathbf{R}^0$, the MSE of the j th variable is approximately $E_j(\sigma_j^0)^2/(E_j + (\sigma_j^0)^2)$, a result that may also be found directly from the asymptotic formula for the scalar case, Equations (40a) and (40b), with $g = 1$, $r^0 = (\sigma^0)^2$, $|\mathbf{F}| \rightarrow 0$, and $r \rightarrow E = \sigma^2/2\mu$. As a result, in the 3D-Var tests with the correct covariance matrix $\mathbf{B} = \mathbf{R}_{\infty+1|\infty}$ the RMSE is expected to

be, after normalizing by $\sqrt{E_j}$ and using the fact that $r_{\mathcal{M}}^o = \alpha\gamma E_{\mathcal{M}}$ and $r_j^o = \alpha E_j$ for $j \in \{\widehat{P}\widehat{V}, \widehat{W}_{1,2}\}$,

$$\frac{\text{RMSE}_j}{\sqrt{E_j}} = \left[\frac{\alpha\gamma}{1 + \alpha\gamma} \right]^{1/2}. \quad (55)$$

In the case of an incorrect (and non-diagonal) covariance matrix $\tilde{\mathbf{B}}$, however, the MSE is

$$\text{MSE} \approx [\tilde{\mathbf{K}}(\mathbf{H}\mathbf{B}\mathbf{H}^* + \mathbf{R}^o) - \mathbf{B}\mathbf{H}^*]\tilde{\mathbf{K}}^* + \mathbf{B} - \tilde{\mathbf{K}}\mathbf{H}\mathbf{B}, \quad (56)$$

where $\tilde{\mathbf{K}} = \tilde{\mathbf{B}}\mathbf{H}^*(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^* + \mathbf{R}^o)^{-1}$. Since the most pronounced differences in analyses were noted for the $\mathcal{M}$ mode, and in order to illustrate the impact of model parameters on the results, we write out explicitly the RMSE for this mode:

$$\frac{\widetilde{\text{RMSE}}_{\mathcal{M}}}{\sqrt{E_{\mathcal{M}}}} = \frac{\text{RMSE}_{\mathcal{M}}}{\sqrt{E_{\mathcal{M}}}} \times \sqrt{1 + \frac{\alpha\gamma B_0[(1 + \alpha)(1 + \alpha\gamma) + \alpha^2 B_0]}{[(1 + \alpha)(1 + \alpha\gamma) + \alpha B_0]^2}}, \quad (57)$$

where

$$B_0 = \tilde{Q}^2 \left[\frac{1}{f^2 R_1^2} \frac{E_{\mathcal{P}\mathcal{V}}}{E_{\mathcal{M}}} \right] \left[\frac{1 + (R_1 \kappa)^2 (E_{\mathcal{W}_j}/E_{\mathcal{P}\mathcal{V}})}{1 + (R_1 \kappa)^2} \right], \quad (58)$$

for $j \in \{1, 2\}$. Thus, $\text{EI} = (\widetilde{\text{RMSE}}_{\mathcal{M}} - \text{RMSE}_{\mathcal{M}})/\text{RMSE}_{\mathcal{M}}$ for the $\mathcal{M}$ mode is

$$\text{EI} = \sqrt{1 + \frac{\alpha\gamma B_0[(1 + \alpha)(1 + \alpha\gamma) + \alpha^2 B_0]}{[(1 + \alpha)(1 + \alpha\gamma) + \alpha B_0]^2}} - 1. \quad (59)$$

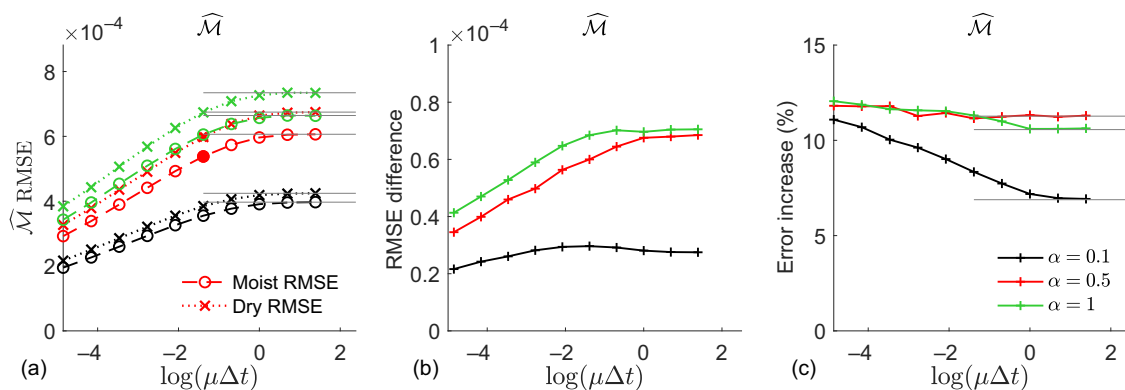


FIGURE 7 (a) Root-mean-square error (RMSE) for $\widehat{\mathcal{M}}$ for various values of $\mu_{\mathcal{M}}\Delta t$ and α . Gray lines indicate values obtained analytically for the large-observation-time limit. (b) Difference between RMSE obtained with dry and moist methods. (c) Error increase resulting from using an erroneous background error covariance matrix.

Figure 8 shows the EI for various values of α and dq_{bg}/dz and two values of γ . As expected, for fixed α , the EI increases monotonically with increasing dq_{bg}/dz , consistent with Figure 5. Also expected is that the EI goes to zero as $\alpha \rightarrow 0$, since having perfect observations makes differences in the background error covariance matrix irrelevant to the analysis. For small α , increasing α results in a higher EI. This increase reaches a maximum at a value of α that may be found from the roots of

$$1 + (2 - B_0)\alpha + B_0\alpha^2 + (-2 + 3B_0 + B_0^2)\alpha^3 - (1 + B_0)\alpha^4 = 0 \quad (60)$$

in the case of $\gamma = 1$. Finally, and perhaps not surprisingly, the higher the value of γ , the larger the EI. Line plots showing the information from Figure 8 for various values of dq_{bg}/dz are shown in Supporting Information Figure S15a,b.

Additional tests were conducted in which only one of the four eigenmodes was observed; specifically, one of $\widehat{\mathcal{M}}$, $\widehat{\mathcal{P}\mathcal{V}}$, or $\widehat{\mathcal{W}}_1$ was observed. For example, in the case where $\widehat{\mathcal{M}}$ was observed, the observation operator $\mathbf{H} = [1 \ 0 \ 0 \ 0]$, and $\mathbf{R}^o$ became a scalar with value given by the corresponding (1, 1) entry in the previous $\mathbf{R}^o$ matrix. The results of these tests are shown in Supporting Information Figures S9–S11. Observing only $\widehat{\mathcal{M}}$ results in changes to the analysis in all mode variables, whereas observing any of the other three results in changes to the analysis of $\widehat{\mathcal{M}}$ only; increases in error of up to 15–20% were seen.

Similarly, PHYS-OBS tests in which only one of the physical variables was observed were also conducted; results are shown in Supporting Information Figures S12–S14. Error increases were largest in the case where only $\hat{\theta}$ was observed, with increases as large as 30% or more seen.

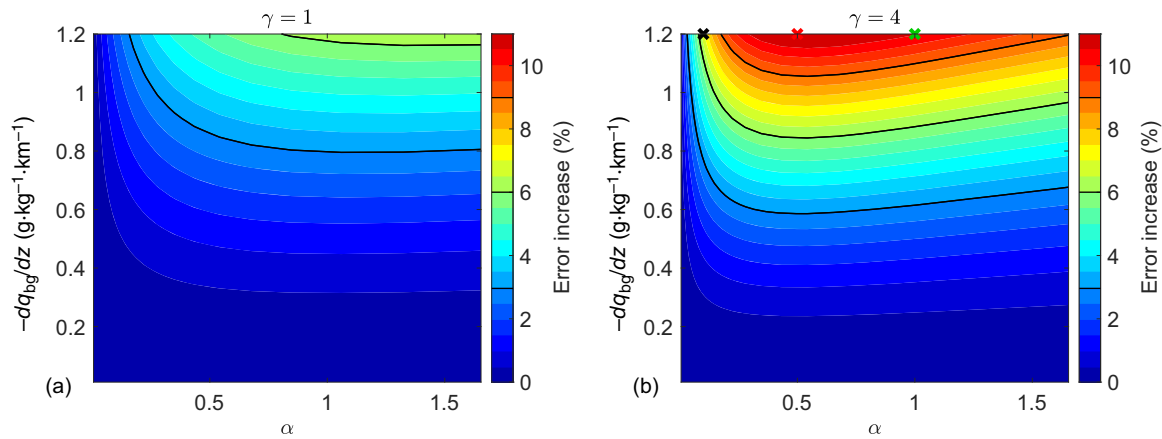


FIGURE 8 (a) Error increase in the infrequent observation limit according to Equation (59) for various values of dq_{bg}/dz and α (relative observation noise) and $\gamma = 1$ (same relative noise for all wave variables). (b) Same as (a) but with $\gamma = 4$ (higher relative observational noise for M). Red, green, and black $\times$'s correspond to parameter values in Figure 7.

4.2 | Unequal dry balanced and unbalanced energy on large or small spatial scales

In addition to the parameter dependencies explored earlier herein (for α , γ , and dq_{bg}/dz), Equation (58) shows that the results could also be impacted by the dry balanced/unbalanced energy ratio E_{W_j}/E_{P_V} , the spatial scale κ , and the balanced energy ratio E_{P_V}/E_M . To probe these dependencies further, we now explore other spatial scales, where the power ratio between balanced and unbalanced modes may not be one.

It is known that the ratio of energy present in balanced and unbalanced modes depends on the spatial scale (e.g., Žagar *et al.*, 2015). On planetary and synoptic scales, balanced modes contain up to 90% or more of the total energy, whereas unbalanced modes may dominate on mesoscales. Figure 9a is a crude reproduction of fig. 7 from Žagar *et al.* (2015), with the percentage of energy in the balanced (unbalanced) modes shown in red (blue). The curves in Figure 9a were used to inform the values of power ratio E_{W_j}/E_{P_V} used for tests with various values of wave number κ . The standard case, representing an even split of dry energy between wave modes and $\widehat{P_V}$, is shown with the shaded circle; two additional cases, which will be referred to as the 10/90 and 80/20 cases, are denoted by open circles and represent the impact of having most energy in balanced and unbalanced modes respectively.

Figure 9b shows the EI in the limit of large time between observations for the combinations of E_{W_j}/E_{P_V} and κ corresponding to Figure 9a; all other parameter values match those of the standard case. The standard case, denoted with a shaded circle, sits at the mesoscale–synoptic scale boundary. On synoptic scales

with smaller κ , the improvement is similar to or less than that of the standard case, with the EI of the state estimate of $\widehat{M}$ between 5% and 10%.

On mesoscales, in Figure 9b, the EI increases significantly, up to 50% or more on the smallest scales considered here. Figure 10 shows the impact of α , γ , and dq_{bg}/dz on these results. Qualitatively, the dependence of EI on each of these parameters is similar to Figure 8, though it is worth pointing out that the color scales are different, with the EI as high as 30% in the 80/20 case. Line plots showing the information from Figure 10 for various values of dq_{bg}/dz are shown in the Supporting Information Figure S15c–f. Error increases of 20–35% are also seen for the 80/20 energy ratio case for shorter observation time; see Supporting Information Figure S7 for these results. We note that although the results here indicate the potential for large improvement on smaller scales where much high-impact weather occurs, the assumptions on which the linear hydrostatic model used here is based break down on smaller scales; implementation of the ideas presented here in a more complicated model able to accurately capture dynamics across a greater range of scales is an important next step for future work.

5 | CONCLUSION

Using a recently formulated description of dynamically balanced versus unbalanced moisture, we have studied the role of moisture balance in a simple model and data assimilation scheme. The dynamically balanced moisture of Smith and Stechmann (2017) and Wetzal *et al.* (2019, 2020) was explored in a simple moist shallow-water system with stochastic forcing, to obtain transparent tests and some analytical formulas. The properties of balanced moisture

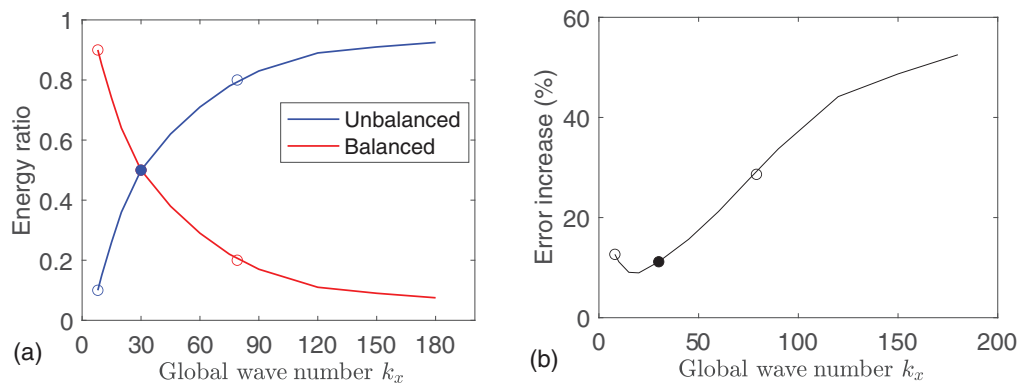
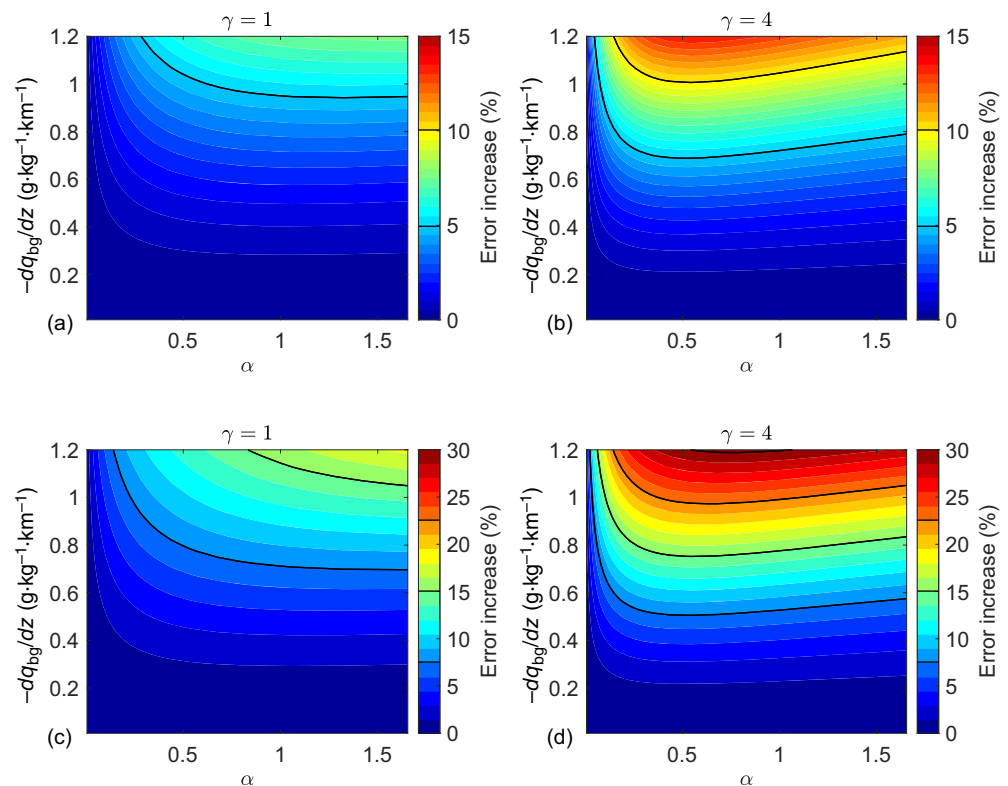


FIGURE 9 (a) Crude reproduction of fig. 7 from Žagar *et al.* (2015), showing the percentage of energy on spatial scales k_x in the unbalanced (blue) and balanced (red) modes. Global wave number k_x is related to model domain wave number k as $k_x \approx 8k$ for model domain size $4,000 \times 4,000 \text{ km}^2$. (b) Error increase with $E_{PV}/E_M = 72$ and E_{W_j}/E_{PV} specified by (a). Circles represent the 10/90, 50/50, and 80/20 unbalanced/balanced energy ratios considered here; closed circle denotes standard case.

FIGURE 10 (a), (b) Same as Figure 8 but for 10/90 dry unbalanced-balanced energy ratio; global wave number $k_x \approx 7$ (corresponding to $k = 1$ on $4,000 \times 4,000 \text{ km}^2$ domain). (c), (d) Same as Figure 8 but for 80/20 dry unbalanced/balanced energy ratio; global wave number $k_x \approx 80$ (corresponding to $k = 11$ on $4,000 \times 4,000 \text{ km}^2$ model domain). Note the different color axes limits for (c) and (d).



are related to some unique or distinctive properties of the moist dynamical evolution equations. In particular, the system's dynamical operator is not symmetric; instead, it is self-adjoint with respect to a non-diagonal inner product, and this inner product is associated with a conserved moist energy.

The main comparison was between using the eigenmodes of the moist dynamics and using an erroneous eigenmode definition that does not properly represent the balanced moisture. The eigenmodes were then used in

a data assimilation scheme akin to 3D-Var, where the eigenmodes are used in diagonalizing the background error covariance matrix. In the numerical experiments, the errors in state estimation of both total moisture and of dynamically balanced moisture were increased by up to 15% or more by using the erroneous approach. These errors are most significant (a) on smaller spatial scales, where a larger portion of the energy is present in the unbalanced modes, (b) with larger vertical gradient of the background moisture state, and (iii) with larger observational noise.

One might expect even larger errors if cloud latent heating and moist convective instability were included. The EI seen here was already 30%, or more, for smaller spatial scales such as mesoscales, although the validity of the linear shallow-water system will break down as nonlinearity and non-hydrostatic effects arise on smaller spatial scales. It would be interesting in the future to investigate nonlinear dynamics, and to apply a nonlinear eigenspace decomposition and balanced–unbalanced decomposition that hold even in the presence of latent heating and nonlinearity due to clouds and phase changes (Remond-Tiedrez *et al.*, 2024).

The results here suggest that a careful accounting of dynamically balanced moisture could potentially have a significant impact on both state estimation and forecasting of moisture in numerical weather prediction. Among many issues that range from fundamental to complex, a longstanding one is the spin-up or spin-down problem for precipitation (Andersson *et al.*, 2005; Carlin *et al.*, 2017; Illari, 1987; Kasahara *et al.*, 1992; Short & Petch, 2022; Sun *et al.*, 2014); this is potentially due to a variety of factors, but perhaps balanced versus unbalanced moisture plays a contributing role as well. Though the model used here was idealized, the results suggest that further testing in more sophisticated models and with a wider variety of data assimilation schemes is warranted. We note that it is possible that some of the difficulties encountered by 3D-Var in accurately accounting for balanced moisture may be partially addressed through the use of ensemble-based methods or hybrid ensemble–variational methods; that is, the errors introduced by a traditional accounting of moisture may be smaller with these more sophisticated methods. However, the use of localization in ensemble methods can introduce problematic issues with balance (Bannister, 2015; Gottwald, 2014; Greybush *et al.*, 2011; Hastermann *et al.*, 2021; Houtekamer & Mitchell, 2005; Kepert, 2009). In any case, the significant improvement obtained by using balanced moisture in 3D-Var, particularly on smaller scales, where a substantial amount of the atmosphere's energy is stored in unbalanced modes, suggests further study in this direction would be useful.

As another direction for possible further study, the balanced M variable could be a candidate for a moisture variable that is suitable for the challenges of data assimilation (e.g., Dee & Da Silva, 2003). For instance, it is a moist quantity that can take either positive or negative values, which is helpful for formulations that require normally distributed prior errors. See Marsico *et al.* (2019) for definitions of M in the setting of the anelastic equations. Also, since M is associated with an independent eigenmode, it may be the moist quantity that is most statistically independent from the vortical/Rossby modes and

inertia–gravity wave modes. Further studies could be helpful for better understanding these properties.

FUNDING INFORMATION

This work was supported by the Simons Foundation, award 854116, and the US NSF DMS, award number 1907667.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

The authors confirm that the data supporting the findings of this study are available within the article and its [Supporting Information](#).

ORCID

H. Reed Ogrosky  <https://orcid.org/0000-0002-5848-834X>

REFERENCES

- Anderson, B.D.O. & Moore, J.B. (1979) *Optimal filtering*. New York: Dover Publications.
- Andersson, E., Bauer, P., Beljaars, A., Chevallier, F., Hólm, E., Janisková, M. et al. (2005) Assimilation and modeling of the atmospheric hydrological cycle in the ECMWF forecasting system. *Bulletin of the American Meteorological Society*, 86(3), 387–402.
- Bannister, R.N. (2015) How is the balance of a forecast ensemble affected by adaptive and nonadaptive localization schemes? *Monthly Weather Review*, 143(9), 3680–3699.
- Bannister, R.N. (2021) Balance conditions in variational data assimilation for a high-resolution forecast model. *Quarterly Journal of the Royal Meteorological Society*, 147(738), 2917–2934.
- Bannister, R.N., Chipilski, H.G. & Martinez-Alvarado, O. (2020) Techniques and challenges in the assimilation of atmospheric water observations for numerical weather prediction towards convective scales. *Quarterly Journal of the Royal Meteorological Society*, 146(726), 1–48.
- Berre, L. (2000) Estimation of synoptic and mesoscale forecast error covariances in a limited-area model. *Monthly Weather Review*, 128(3), 644–667.
- Bretherton, C.S. & Smolarkiewicz, P.K. (1989) Gravity waves, compensating subsidence and detrainment around cumulus clouds. *Journal of the Atmospheric Sciences*, 46(6), 740–759.
- Bube, K.P. & Ghil, M. (1981) Assimilation of synoptic data and the initialization problem. In: Bengtsson, L., Ghil, M. & Källén, E. (Eds.) *Dynamic meteorology: Data assimilation methods*. New York: Springer, pp. 111–138.
- Carlin, J.T., Gao, J., Snyder, J.C. & Ryzhkov, A.V. (2017) Assimilation of z_{dr} columns for improving the spinup and forecast of convective storms in storm-scale models: proof-of-concept experiments. *Monthly Weather Review*, 145(12), 5033–5057.
- Chen, N. & Majda, A.J. (2016) Filtering the stochastic skeleton model for the Madden–Julian oscillation. *Monthly Weather Review*, 144(2), 501–527.

- Chen, Y., Rizvi, S.R.H., Huang, X.-Y., Min, J. & Zhang, X. (2013) Balance characteristics of multivariate background error covariances and their impact on analyses and forecasts in tropical and Arctic regions. *Meteorology and Atmospheric Physics*, 121, 79–98.
- Chen, Y. & Stechmann, S.N. (2019) Multi-model communication and data assimilation for mitigating model error and improving forecasts. *Chinese Annals of Mathematics, Series B*, 40, 689–720.
- Daley, R. (1993) *Atmospheric data analysis*, Vol. 2. Cambridge: Cambridge University Press.
- Dee, D.P. & Da Silva, A.M. (2003) The choice of variable for atmospheric moisture analysis. *Monthly Weather Review*, 131(1), 155–171.
- DelSole, T. (2004) Stochastic models of quasigeostrophic turbulence. *Surveys in Geophysics*, 25(2), 107–149. Available from: <https://doi.org/10.1023/B:GEOP.0000028160.75549.0d>
- Derber, J. & Bouttier, F. (1999) A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus A*, 51(2), 195–221.
- Field, P.R. & Wood, R. (2007) Precipitation and cloud structure in midlatitude cyclones. *Journal of Climate*, 20, 233–254.
- Goodliff, M., Amezcua, J. & Van Leeuwen, P.J. (2017) A weak-constraint 4D EnsembleVar. Part II: experiments with larger models. *Tellus A: Dynamic Meteorology and Oceanography*, 69(1), 1271565.
- Gottwald, G.A. (2014) Controlling balance in an ensemble Kalman filter. *Nonlinear Processes in Geophysics*, 21(2), 417–426.
- Greybush, S.J., Kalnay, E., Miyoshi, T., Ide, K. & Hunt, B.R. (2011) Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review*, 139(2), 511–522.
- Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H. et al. (2018) Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society*, 144(713), 1218–1256.
- Harlim, J. & Majda, A.J. (2008) Filtering nonlinear dynamical systems with linear stochastic models. *Nonlinearity*, 21(6), 1281–1306.
- Harlim, J. & Majda, A.J. (2013) Test models for filtering and prediction of moisture-coupled tropical waves. *Quarterly Journal of the Royal Meteorological Society*, 139(670), 119–136.
- Hastermann, G., Reinhardt, M., Klein, R. & Reich, S. (2021) Balanced data assimilation for highly oscillatory mechanical systems. *Communications in Applied Mathematics and Computational Science*, 16(1), 119–154.
- Horn, R.A. & Johnson, C.R. (2012) *Matrix analysis*. New York, NY: Cambridge University Press.
- Hottovy, S. & Stechmann, S.N. (2015) A spatiotemporal stochastic model for tropical precipitation and water vapor dynamics. *Journal of the Atmospheric Sciences*, 72, 4721–4738. Available from: <https://doi.org/10.1175/JAS-D-15-0119.1>
- Houtekamer, P.L. & Mitchell, H.L. (2005) Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3269–3289.
- Huang, T., Stechmann, S.N. & Torchinsky, J.L. (2022) Framework for idealized climate simulations with spatiotemporal stochastic clouds and planetary-scale circulations. *Physical Review Fluids*, 7(1), 010502.
- Ide, K., Courtier, P., Ghil, M. & Lorenc, A.C. (1997) Unified notation for data assimilation: operational, sequential, 714 and variational. *Journal of the Meteorological Society of Japan*, 75(1B), 181–189.
- Illari, L. (1987) The spin up problem. ECMWF Tech. Memo. No. 137.
- Ingleby, N.B., Lorenc, A.C., Ngan, K., Rawlins, F. & Jackson, D.R. (2013) Improved variational analyses using a nonlinear humidity control variable. *Quarterly Journal of the Royal Meteorological Society*, 139(676), 1875–1887.
- Janjić, T., McLaughlin, D., Cohn, S.E. & Verlaan, M. (2014) Conservation of mass and preservation of positivity with ensemble-type Kalman filter algorithms. *Monthly Weather Review*, 142(2), 755–773.
- Janjić, T., Ruckstuhl, Y. & Toint, P.L. (2021) A data assimilation algorithm for predicting rain. *Quarterly Journal of the Royal Meteorological Society*, 147(736), 1949–1963.
- Kalnay, E. (2003) *Atmospheric modeling, data assimilation and predictability*. Cambridge: Cambridge University Press.
- Kasahara, A. & Puri, K. (1981) Spectral representation of three-dimensional global data by expansion in normal mode functions. *Monthly Weather Review*, 109(1), 37–51.
- Kasahara, A., Mizzi, A.P. & Donner, L.J. (1992) Impact of cumulus initialization on the spinup of precipitation forecasts in the tropics. *Monthly Weather Review*, 120(7), 1360–1380.
- Kent, T., Bokhove, O. & Tobias, S. (2017) Dynamics of an idealized fluid model for investigating convective-scale data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 69(1), 1369332.
- Keptert, J.D. (2009) Covariance localisation and balance in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135(642), 1157–1176.
- Khouider, B. & Majda, A.J. (2006) Multicloud convective parameterizations with crude vertical structure. *Theoretical and Computational Fluid Dynamics*, 20, 351–375.
- Kooloth, P., Smith, L.M. & Stechmann, S.N. (2023) Hamilton's principle with phase changes and conservation principles for moist potential vorticity. *Quarterly Journal of the Royal Meteorological Society*, 149(752), 1056–1072. Available from: <https://doi.org/10.1002/qj.4454>
- Liu, C. & Moncrieff, M.W. (2004) Effects of convectively generated gravity waves and rotation on the organization of convection. *Journal of the Atmospheric Sciences*, 61(17), 2218–2227.
- Lorenc, A.C. (1981) A global three-dimensional multivariate statistical interpolation scheme. *Monthly Weather Review*, 109(4), 701–721.
- Majda, A.J. (2003) *Introduction to PDEs and waves for the atmosphere and ocean*. Courant Lecture Notes in Mathematics, Vol. 9. Providence: American Mathematical Society.
- Majda, A.J. & Harlim, J. (2012) *Filtering complex turbulent systems*. Cambridge: Cambridge University Press.
- Mapes, B. (1993) Gregarious tropical convection. *Journal of the Atmospheric Sciences*, 50(13), 2026–2037.
- Marsico, D.H., Smith, L.M. & Stechmann, S.N. (2019) Energy decompositions for moist Boussinesq and anelastic equations with phase changes. *Journal of the Atmospheric Sciences*, 76, 3569–3587. Available from: <https://doi.org/10.1175/JAS-D-19-0080.1>
- Michel, Y., Auligné, T. & Montmerle, T. (2011) Heterogeneous convective-scale background error covariances with the inclusion of hydrometeor variables. *Monthly Weather Review*, 139(9), 2994–3015.
- Montmerle, T. (2012) Optimization of the assimilation of radar data at the convective scale using specific background error

- covariances in precipitation. *Monthly Weather Review*, 140(11), 3495–3506.
- Montmerle, T. & Berre, L. (2010) Diagnosis and formulation of heterogeneous background-error covariances at the mesoscale. *Quarterly Journal of the Royal Meteorological Society*, 136(651), 1408–1420.
- Mou, C., Smith, L.M. & Chen, N. (2023) Combining stochastic parameterized reduced-order models with machine learning for data assimilation and uncertainty quantification with partial observations. *Journal of Advances in Modeling Earth Systems*, 15(10), e2022MS003597.
- Platzman, G.W. (1967) A retrospective view of Richardson's book on weather prediction. *Bulletin of the American Meteorological Society*, 48(8), 514–551.
- Remond-Tiedrez, A., Smith, L.M. & Stechmann, S.N. (2024) Non-linear eigenspace decomposition for a moist atmosphere with clouds. (in preparation).
- Richardson, L.F. (1922) *Weather prediction by numerical process*. Cambridge: University Press.
- Robert, S. & Künsch, H.R. (2017) Localizing the ensemble Kalman particle filter. *Tellus A: Dynamic Meteorology and Oceanography*, 69(1), 1282016.
- Short, C.J. & Petch, J. (2022) Reducing the spin-up of a regional NWP system without data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 148(745), 1623–1643.
- Smith, L.M. & Stechmann, S.N. (2017) Precipitating quasigeostrophic equations and potential vorticity inversion with phase changes. *Journal of the Atmospheric Sciences*, 74, 3285–3303. Available from: <https://doi.org/10.1175/JAS-D-17-0023.1>
- Stechmann, S.N. & Hottovy, S. (2017) Unified spectrum of tropical rainfall and waves in a simple stochastic model. *Geophysical Research Letters*, 44, 10,713–10,724. Available from: <https://doi.org/10.1002/2017GL075754>
- Stechmann, S.N. & Majda, A.J. (2009) Gravity waves in shear and implications for organized convection. *Journal of the Atmospheric Sciences*, 66, 2579–2599.
- Stechmann, S.N. & Majda, A.J. (2015) Identifying the skeleton of the madden–Julian oscillation in observational data. *Monthly Weather Review*, 143, 395–416. Available from: <https://doi.org/10.1175/MWR-D-14-00169.1>
- Stechmann, S.N., Majda, A.J. & Khouider, B. (2008) Nonlinear dynamics of hydrostatic internal gravity waves. *Theoretical and Computational Fluid Dynamics*, 22, 407–432.
- Sun, J., Xue, M., Wilson, J.W., Zawadzki, I., Ballard, S.P., Onvlee-Hoimeyer, J. et al. (2014) Use of NWP for nowcasting convective precipitation: recent progress and challenges. *Bulletin of the American Meteorological Society*, 95(3), 409–426.
- Talagrand, O. (1981) A study of the dynamics of four-dimensional data assimilation. *Tellus*, 33(1), 43–60.
- Wetzel, A.N., Smith, L.M., Stechmann, S.N. & Martin, J.E. (2019) Balanced and unbalanced components of moist atmospheric flows with phase changes. *Chinese Annals of Mathematics, Series B*, 40, 1005–1038.
- Wetzel, A.N., Smith, L.M., Stechmann, S.N., Martin, J.E. & Zhang, Y. (2020) Potential vorticity and balanced and unbalanced moisture. *Journal of the Atmospheric Sciences*, 77, 1913–1931. Available from: <https://doi.org/10.1175/JAS-D-19-0311.1>
- Würsch, M. & Craig, G.C. (2014) A simple dynamical model of cumulus convection for data assimilation research. *Meteorologische Zeitschrift*, 5, 483–490.
- Xiao, Q. & Sun, J. (2007) Multiple-radar data assimilation and short-range quantitative precipitation forecasting of a squall line observed during IHOP 2002. *Monthly Weather Review*, 135(10), 3381–3404.
- Žagar, N., Akira Kasahara, K., Terasaki, J.T. & Tanaka, H. (2015) Normal-mode function representation of global 3-d data sets: open-access software for the atmospheric research community. *Geoscientific Model Development*, 8(4), 1169–1195.
- Zagar, N., Gustafsson, N. & Kallen, E. (2004a) Variational data assimilation in the tropics: the impact of a background-error constraint. *Quarterly Journal of the Royal Meteorological Society*, 130(596), 103–125.
- Zagar, N., Gustafsson, N. & Kallen, E. (2004b) Dynamical response of equatorial waves in four-dimensional variational data assimilation. *Tellus A*, 56, 29–46.
- Žagar, N. (2012) Multivariate data assimilation in the tropics by using equatorial waves. *Pure and Applied Geophysics*, 169, 367–379.
- Zaplotnik, Ž., Žagar, N. & Gustafsson, N. (2018) An intermediate-complexity model for four-dimensional variational data assimilation including moist processes. *Quarterly Journal of the Royal Meteorological Society*, 144(715), 1772–1787.
- Zhang, Y., Smith, L.M. & Stechmann, S.N. (2021a) Effects of clouds and phase changes on fast-wave averaging: a numerical assessment. *Journal of Fluid Mechanics*, 920, A49–1–A49–40. Available from: <https://doi.org/10.1017/jfm.2021.427>
- Zhang, Y., Smith, L.M. & Stechmann, S.N. (2021b) Fast-wave averaging with phase changes: asymptotics and application to moist atmospheric dynamics. *Journal of Nonlinear Science*, 31(38), 1–46. Available from: <https://doi.org/10.1007/s00332-021-09697-2>
- Zhang, Y., Smith, L.M. & Stechmann, S.N. (2022) Convergence to precipitating quasi-geostrophic equations with phase changes: asymptotics and numerical assessment. *Philosophical Transactions of the Royal Society A*, 380(2226), 20210030. Available from: <https://doi.org/10.1098/rsta.2021.0030>
- Zhu, J. & Bannister, R.N. (2023) The hydro-ABC model (version 2.0): a simplified convective-scale model with moist dynamics. *Geoscientific Model Development*, 16(21), 6067–6085.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ogrosky, H.R., Wetzel, A.N., Smith, L.M. & Stechmann, S.N. (2025) Data assimilation errors due to incomplete accounting of balanced versus unbalanced moisture. *Quarterly Journal of the Royal Meteorological Society*, 151:e4942. Available from: <https://doi.org/10.1002/qj.4942>