
SUPPLEMENTARY MATERIAL

Supplementary Material for ‘Data assimilation errors due to incomplete accounting of balanced versus unbalanced moisture’

H. Reed Ogrosky¹ | Alfredo N. Wetzel² | Leslie M. Smith^{2,3} | Samuel N. Stechmann^{2,3}

¹Department of Mathematics and Applied Mathematics, Virginia Commonwealth University, Richmond, VA, 23284, USA

²Department of Mathematics, University of Wisconsin-Madison, Madison, WI, 53706, USA

³Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, WI, 53706, USA

Correspondence

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

Funding information

Simons Foundation, Award 854116;
US NSF DMS, Award Number 1907667

1 | EIG-OBS SIMULATIONS IN u -SPACE

In the main text, figure 4 shows the final 10 days of 3D-Var assimilation cycles using the standard parameter values; all results there are shown in terms of mode variables $\hat{\varphi}$. Here, figure S1 shows the same data, but in terms of the physical (Fourier) variables $\hat{u}$. The rapid oscillations associated with the unbalanced modes $\hat{W}_{1,2}$ can be seen in all four variables, as can the slow oscillations associated with the balanced modes $\mathcal{P}^{\hat{v}}\mathcal{V}$ and $\hat{\mathcal{M}}$.

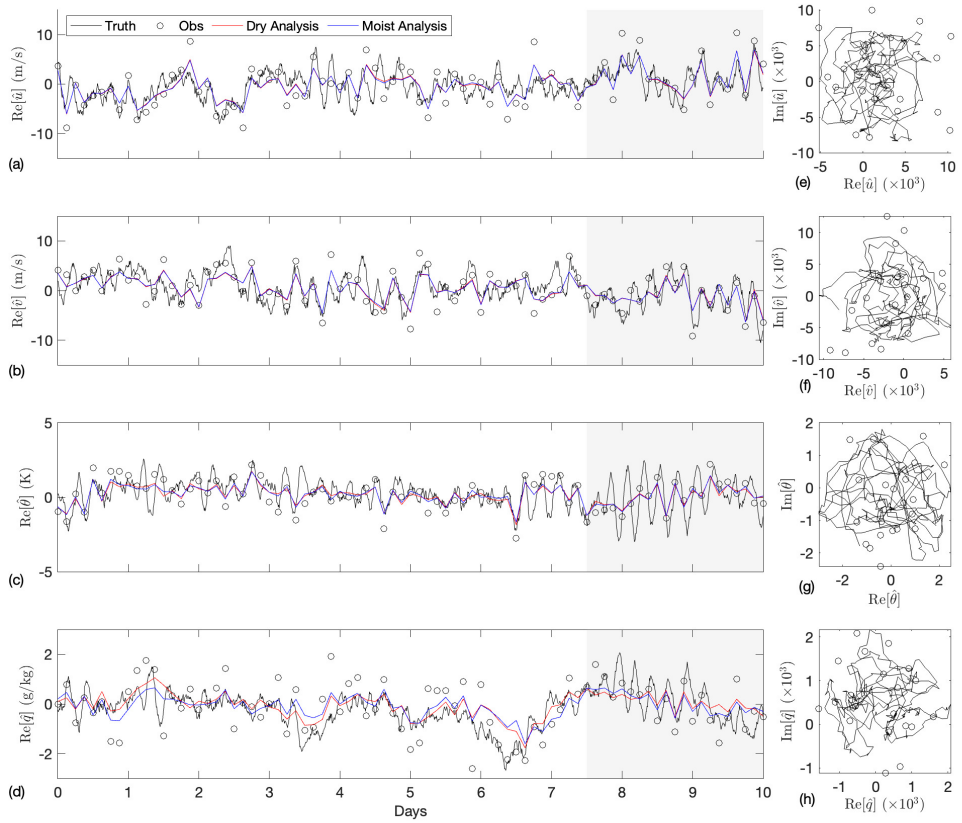


FIGURE S1 Same as figure 4 in the main text but shown in $\mathbf{u}$ space. Final 10 days from 3D-Var with standard parameter values in table 1 in the main text. (a)-(d) Real part of the true signal (solid black), observations (black circles), moist analysis (blue line) and dry analysis (red line) for physical variables $\hat{u}$, $\hat{v}$, $\hat{\theta}$, and $\hat{q}$. (e)-(h) Real and imaginary parts of true signal and observations for the last 2.5 days (shaded region of (a)-(d)).

2 | PHYS-OBS RESULTS

In this section, the results of the PHYS-OBS tests referred to in the main text are presented. These tests are identical to the EIG-OBS tests discussed in detail in the main text, with the exception that the observations were constructed by adding independent noise to the truth signal in $\hat{\mathbf{u}}$ space instead of $\hat{\varphi}$ space. The matrices $\mathbf{H}$ and $\mathbf{R}_{m+1}^o$ used in these tests are defined in the main text.

Figures S2 and S3 here show the final 10 days of 3D-Var assimilation cycles in the PHYS-OBS tests using the standard parameter values; figure S2 shows the results in terms of mode variables $\hat{\varphi}$ while figure S3 shows results in terms of physical variables $\hat{\mathbf{u}}$. The primary features of these figures are similar to the corresponding EIG-OBS results; the unbalanced (balanced) modes exhibit rapid (slow) oscillations in figure S2, while evidence of both balanced and unbalanced modes can be seen in the physical variables in figure S3.

The RMSE for the PHYS-OBS tests with standard parameter values and various values of dq_{bg}/dz are shown in figure S4. As with the EIG-OBS tests, the error is most pronounced in the $\hat{\mathcal{M}}$ mode ($\hat{q}$ in physical space). The error increase is slightly higher than in the EIG-OBS tests, with increase as high as 15% seen in the state estimation of $\hat{\mathcal{M}}$. The error increase seen in estimates of $\hat{\theta}$ is up to 3%.

As with the EIG-OBS tests, the error increase may be found analytically in the limiting case of large time between observations. Due to the non-diagonal nature of the matrices in the PHYS-OBS tests, algebraic formulas provide less insight than in the EIG-OBS case, and the error increase is simply shown in Figure S5 for the standard parameter values and various values of α , γ , and dq_{bg}/dz . The results are quite similar to those of figure 5 in the main text, though once again, there is a very slight increase in the EI seen in the PHYS-OBS tests relative to the EIG-OBS tests, consistent with figure S4.

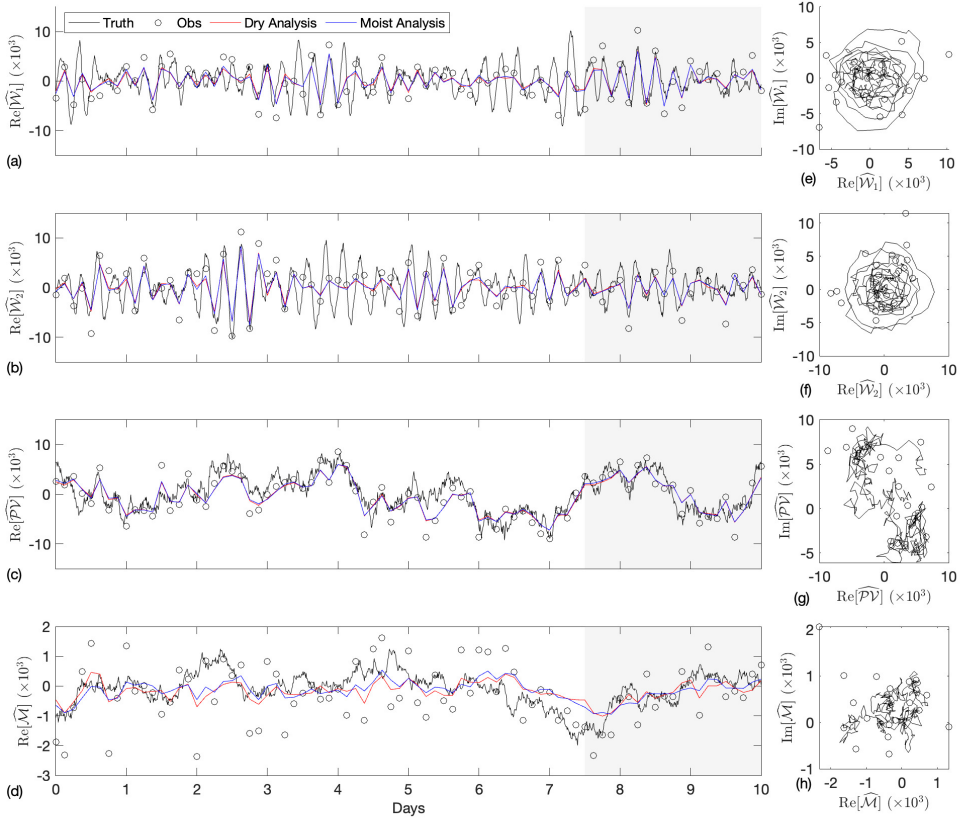


FIGURE S2 Same as figure 4 in main text, but here for the PHYS-OBS tests. Final 10 days from 3D-Var PHYS-OBS tests with standard parameter values in table 1 in the main text. (a)-(d) Real part of the true signal (solid black line), observations (black circles), moist analysis (blue line) and dry analysis (red line) for wave variables $\hat{\mathcal{M}}$, $\hat{\mathcal{P}}^{\hat{V}}$, $\hat{\mathcal{W}}_1$, and $\hat{\mathcal{W}}_2$. (e)-(h) Real and imaginary parts of true signal and observations for the last 2.5 days (shaded region of (a)-(d)).

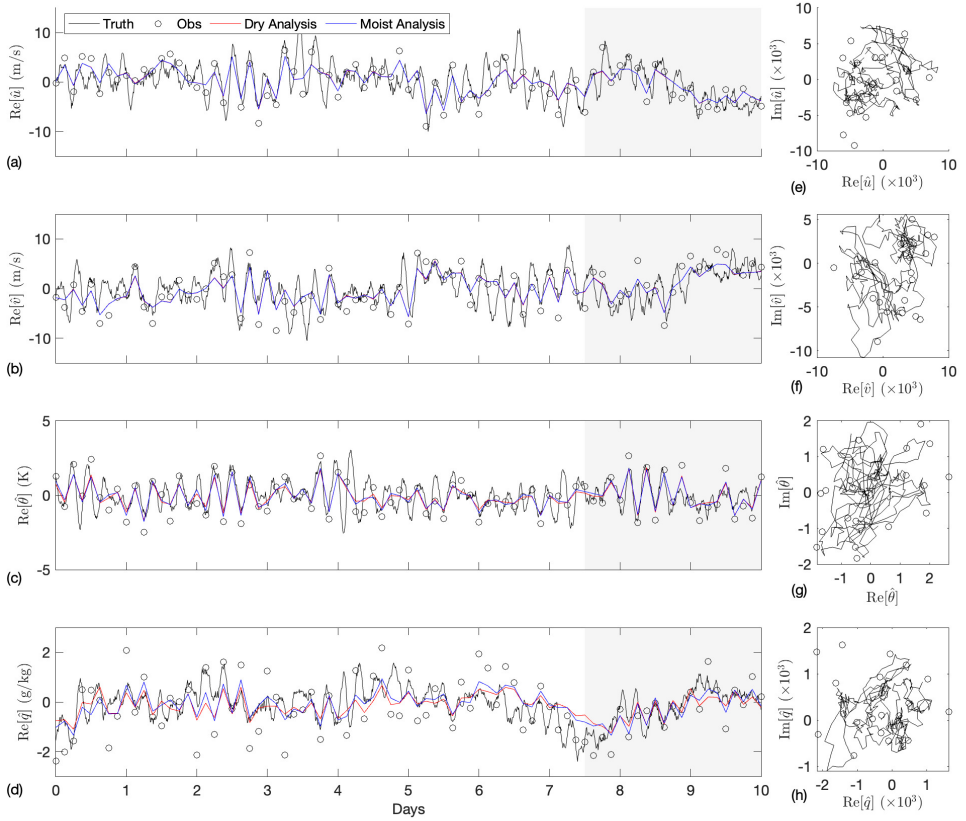


FIGURE S3 Same as figure S1 in supplementary materials, but here for the PHYS-OBS tests. Final 10 days from 3D-Var PHYS-OBS tests with standard parameter values in table 1 in the main text. (a)-(d) Real part of the true signal (solid black line), observations (black circles), moist analysis (blue line) and dry analysis (red line) for physical variables $\hat{u}$, $\hat{v}$, $\hat{\theta}$, and $\hat{q}$. (e)-(h) Real and imaginary parts of true signal and observations for the last 2.5 days (shaded region of (a)-(d)).

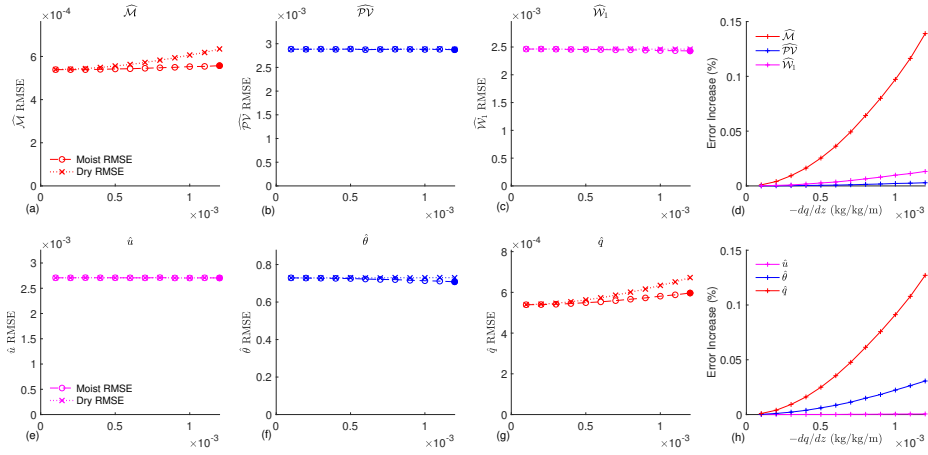


FIGURE S4 Same as figure 5 in main text, but here for the PHYS-OBS tests. (a)-(c) RMSE for mode variables for various values of dq_{bg}/dz . $\hat{W}_2$ results are nearly identical to those of $\hat{W}_1$ and are omitted. (d) Error Increase for mode variables due to using $\hat{\mathbf{B}}$ instead of $\mathbf{B}$ (see main text for definitions). (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.

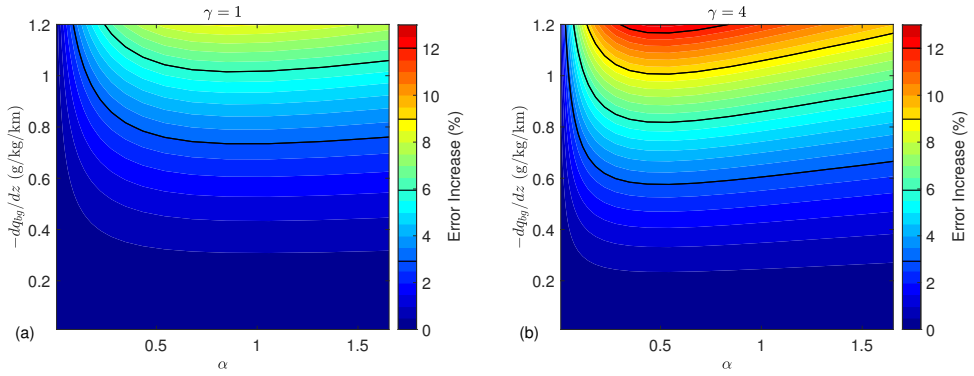


FIGURE S5 Same as figure 8 in main text, but here for the PHYS-OBS tests. (a) Error Increase EI in the low-frequency observation limit for various values of dq_{bg}/dz and α (relative observation noise) and $\gamma = 1$ (same relative noise for all wave variables). (b) Same as (a) but with $\gamma = 4$ (higher relative observational noise for M).

3 | ADDITIONAL EIG-OBS RESULTS

Figure S6 is identical to figure 7 in the main text except that here $dq_{bg}/dz = -0.6 \text{ g/kg/km}$. As may be expected from figure 5 in the main text, the RMSE difference and error increase are smaller with this smaller magnitude of dq_{bg}/dz .

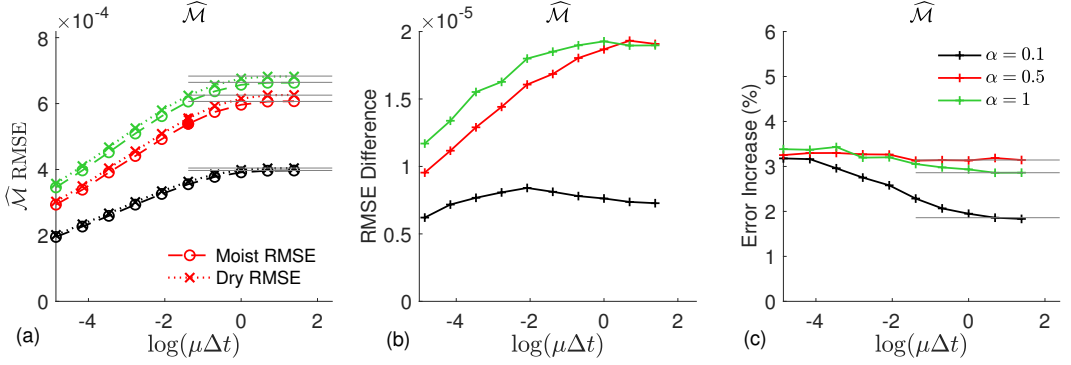


FIGURE S6 Same as figure 7 in main text, but here for $dq_{bg}/dz = -0.6 \text{ g/kg/km}$.

Figure S7 is identical to figure 7 in the main text but for small spatial scales such that the ratio of unbalanced to balanced energy is 80/20. As in the 50/50 energy ratio case, the RMSE increases with increasing α , and the dry analysis is outperformed by the moist analysis. Figure S7(b) shows the difference between RMSE for the two methods in panel (a); figure S7(c) shows that error increases of 20-35% are seen for this small-scale case.

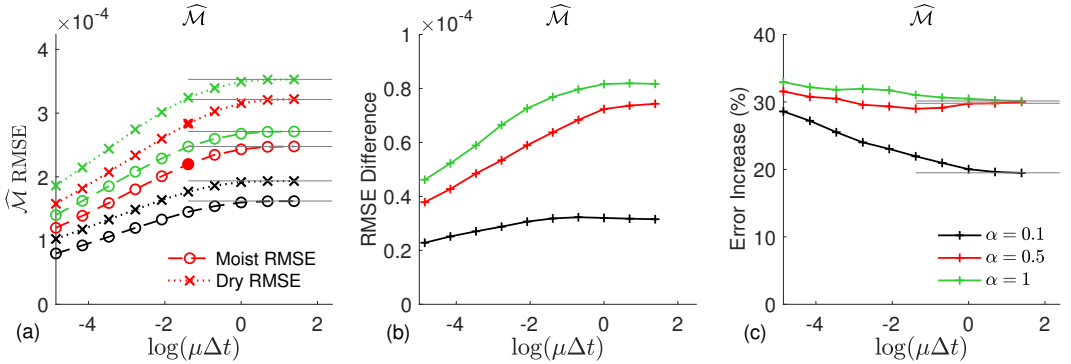


FIGURE S7 Same as figure 7 in main text, but here for 80/20 unbalanced-balanced energy ratio, with global wavenumber $k_x \approx 80$ (corresponding to $k = 11$ on $4000 \times 4000 \text{ km}$ model domain). Stochastic parameters chosen so that $E_{\mathcal{M}} = 9.1 \times 10^{-8}$, $E_{\mathcal{P}\gamma} = 6.6 \times 10^{-6}$, $E_{\mathcal{W}_{1,2}} = 1.3 \times 10^{-5}$.

Figure S8 shows the real part of the $\widehat{\mathcal{M}}$ mode over a 40-day period of a simulation. Times when the moist analysis (in the complex plane) is closer to the truth are depicted with grey shading. The moist analysis is closer to the truth than the dry analysis roughly two-thirds of the time.

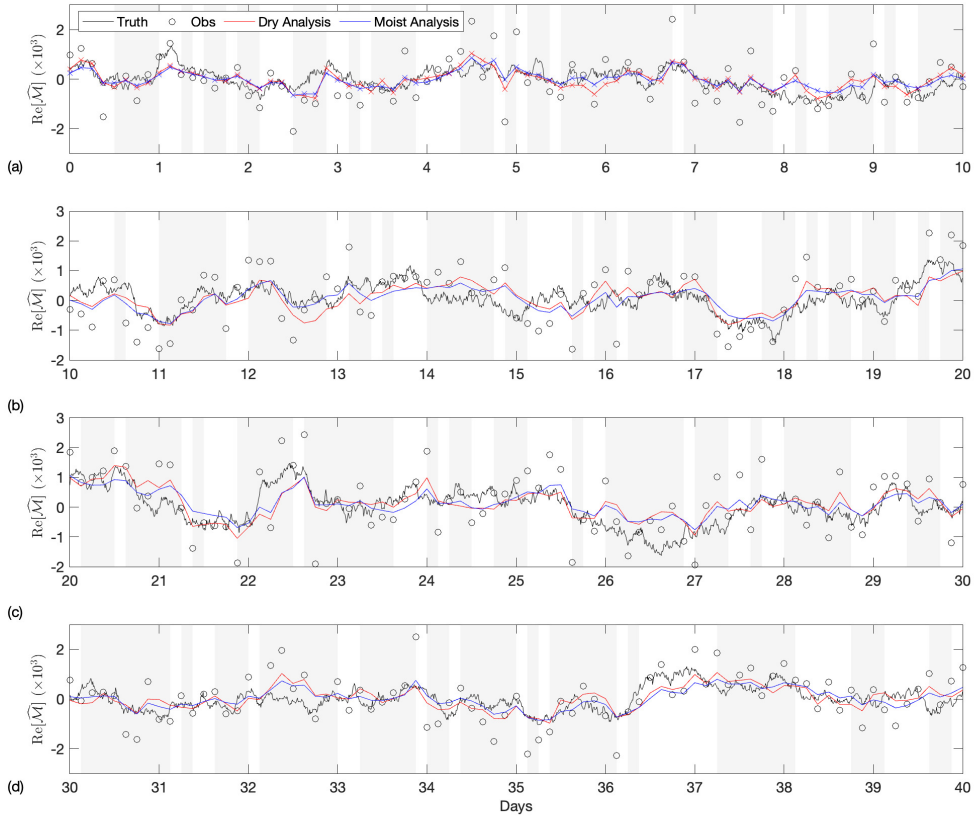


FIGURE S8 Final 40 days from 3D-Var with standard parameter values in table 1. (a)-(d) Real part of the true signal (solid black), observations (black circles), moist analysis (equation (41), blue line) and dry analysis (equation (41), except $\mathbf{B}$ replaced by $\tilde{\mathbf{B}}$, red line) for eigenmode variable $\mathcal{M}_1$. Grey shading depicts times when moist analysis is closer to the truth in the complex plane.

4 | SINGLE-VARIABLE OBSERVATION RESULTS

Tests were also conducted in which only one of the eigenmodes was observed; specifically one of $\hat{\mathcal{M}}$, $\hat{\mathcal{P}}\hat{\mathcal{V}}$, or $\hat{\mathcal{W}}_1$ was observed. For example, in the case where $\hat{\mathcal{M}}$ 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 figures S9-S11. Note that when either $\hat{\mathcal{P}}\hat{\mathcal{V}}$ or $\hat{\mathcal{W}}_1$ are observed, there can be a significant difference between the moist and dry analysis for moisture variable $\hat{\mathcal{M}}$, but no difference between the moist and dry analyses for $\hat{\mathcal{P}}\hat{\mathcal{V}}$ or $\hat{\mathcal{W}}_1$. On the other hand, if the only variable observed is $\hat{\mathcal{M}}$, then all of the eigenmodes are impacted in the analysis.

Similarly, figures S12-S14 show the results of observing only one physical variable. Observing $\hat{q}$ results in differing analysis for all variables, with greatest impact on the dry variables, while observations of $\hat{u}$ or $\hat{\theta}$ only affect only $\hat{\mathcal{M}}$ (or $\hat{q}$), with very large impacts seen with $\hat{\theta}$ observations (up to 40% error increase).

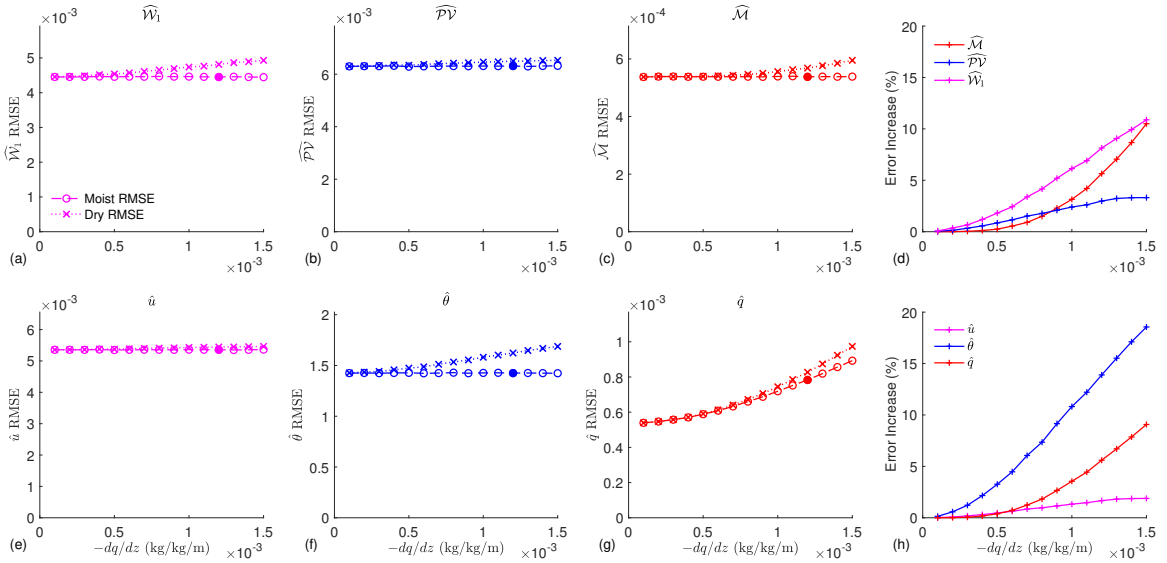


FIGURE S9 Same as figure 5 in manuscript, but with observations of $\hat{\mathcal{M}}$ only.

Figure S15 shows slices of the contour plots in figures 8 and 10 of the main text for various values of dq_{bg}/dz .

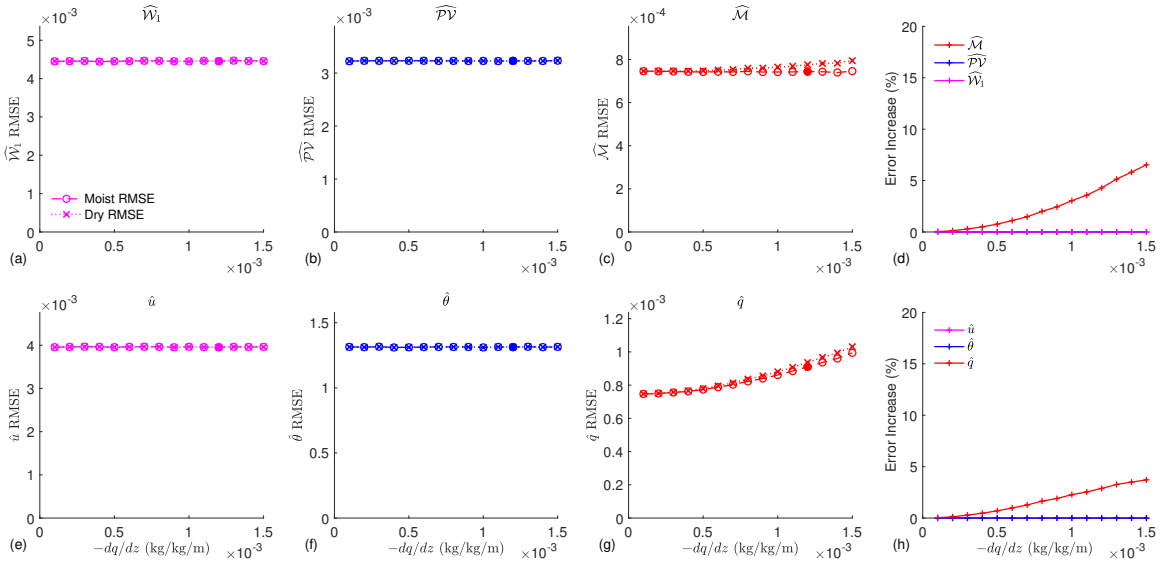


FIGURE S10 Same as figure 5 in manuscript, but with observations of $\mathcal{P}^\gamma \mathcal{V}$ only.

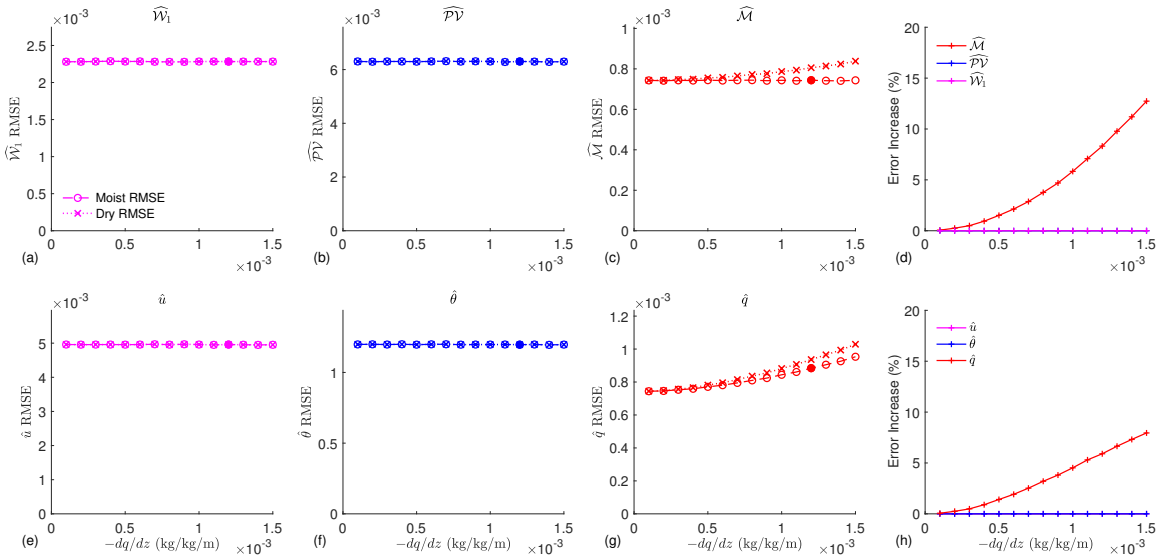


FIGURE S11 Same as figure 5 in manuscript, but with observations of $\mathcal{W}_1^\gamma$ only.

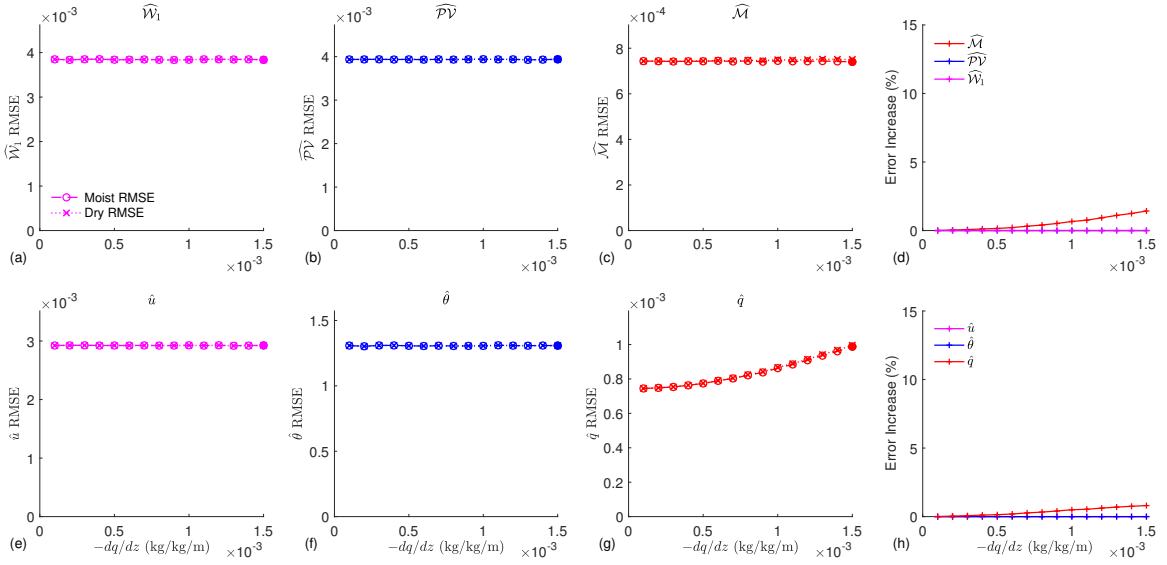


FIGURE S12 Same as figure 5 in manuscript, but with observations of $\hat{u}$ only.

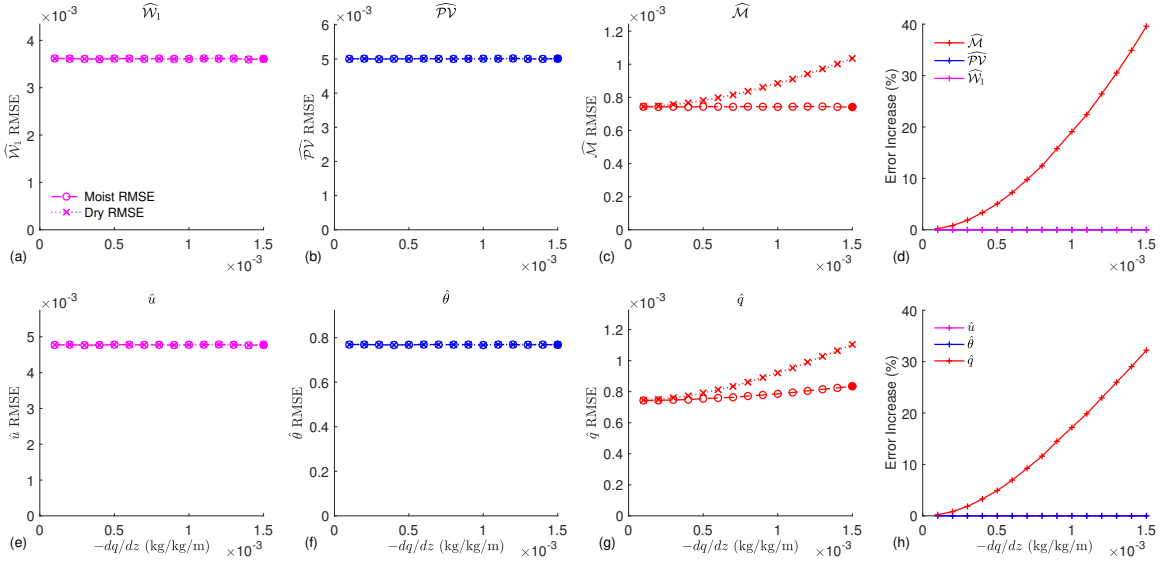


FIGURE S13 Same as figure 5 in manuscript, but with observations of $\hat{\theta}$ only.

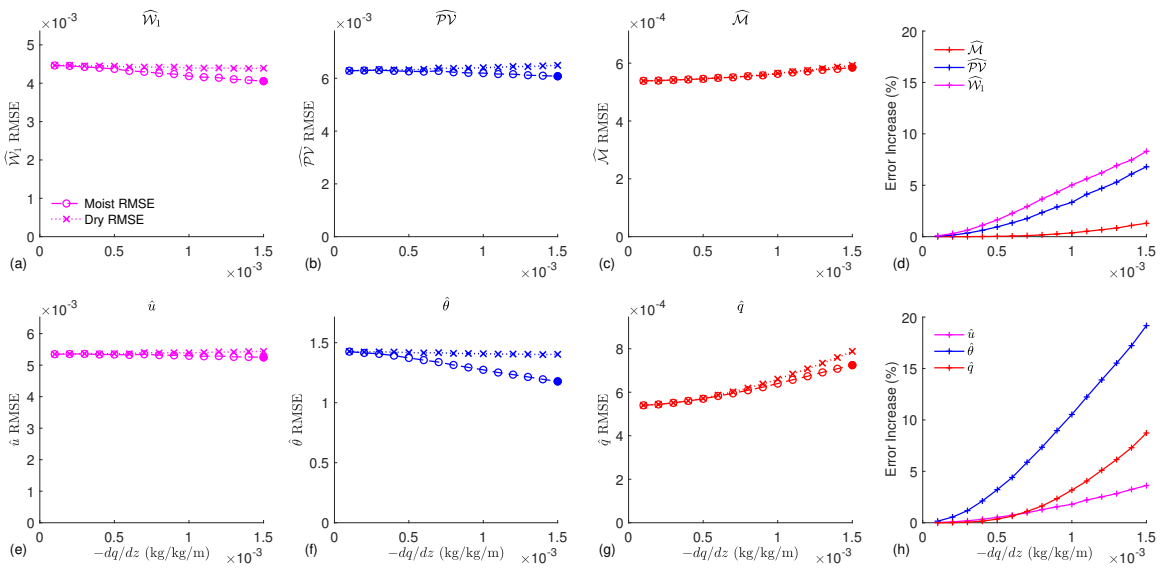


FIGURE S14 Same as figure 5 in manuscript, but with observations of $\hat{q}$ only.

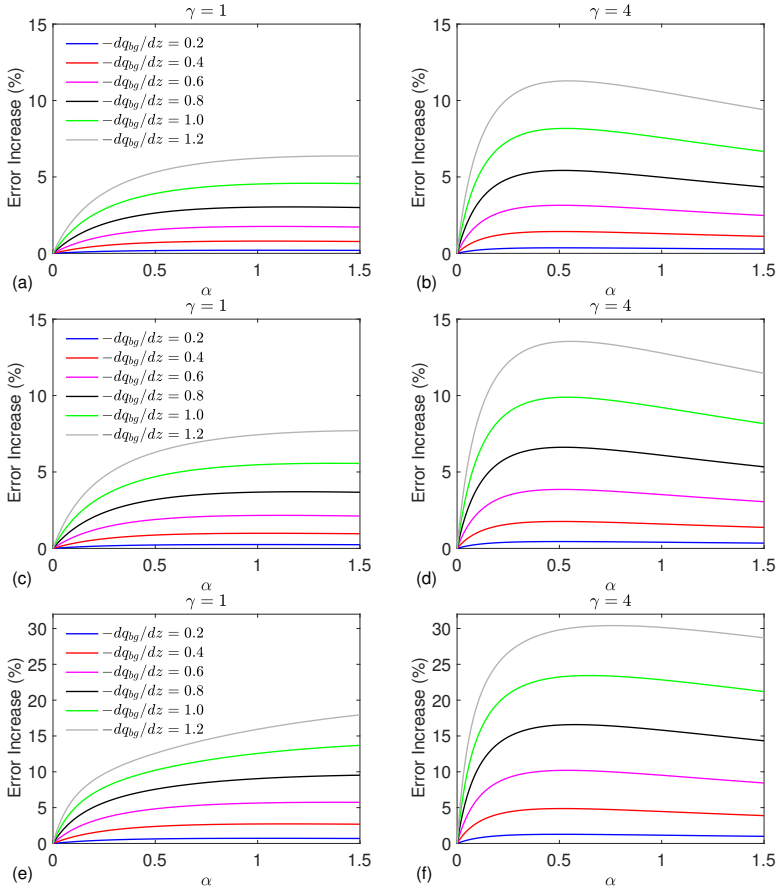


FIGURE S15 (a)-(b) Same information as figure 8 but with slices of constant dq_{bg}/dz . (c)-(f) Same information as figure 10 but with slices of constant dq_{bg}/dz .

5 | DERIVATION OF FORMULA FOR MEAN SQUARE ERROR

In this section, some details are given for the derivation of the analytical formula for both the moist RMSE and dry RMSE in the large-observation-time limit. In this limit, the model only contributes knowledge of the climatological covariance to the assimilation process; the prior is thus very close to zero and can be neglected. For example, using the Kalman filter in this limit, the matrix $A \rightarrow 0$ so that the prior mean and covariance are

$$\tilde{\varphi}_{m+1|m} \approx 0, \quad (1a)$$

$$\mathbf{R}_{m+1|m} \approx \mathbf{R}, \quad (1b)$$

with $\mathbf{R}$ a diagonal matrix with the j th entry along the diagonal given by $r_j = \sigma_j^2 / 2\mu_j$. As a result, the Kalman gain is given by

$$\mathbf{K}_{m+1} \approx \mathbf{R}\mathbf{H}^*(\mathbf{H}\mathbf{R}\mathbf{H}^* + \mathbf{R}^o)^{-1}, \quad (2)$$

with the approximate posterior mean and covariance given by

$$\tilde{\varphi}_{m+1|m+1} \approx \mathbf{K}_{m+1}\hat{\psi}_{m+1} = \mathbf{K}_{m+1}(\mathbf{G}\tilde{\varphi}_{m+1} + \sigma_{m+1}^o), \quad (3a)$$

$$\mathbf{R}_{m+1|m+1} \approx (\mathbf{I} - \mathbf{K}_{m+1}\mathbf{H})\mathbf{R}, \quad (3b)$$

producing the MSE given in equation (54) in the main text; for the j th variable, this is approximately $r_j r_j^o / (r_j + r_j^o)$. Note that in the asymptotic Kalman filter formulas, this is equivalent to taking $\mathbf{F} \rightarrow 0$.

For the 3D-Var assimilation cycles, the background error covariance matrix is specified to be either (i) $\mathbf{B} = \mathbf{R}$ (moist case) or (ii) $\tilde{\mathbf{B}} = \mathbf{T}_{kl}\tilde{\mathbf{T}}_{kl}^{-1}\mathbf{R}(\mathbf{T}_{kl}\tilde{\mathbf{T}}_{kl}^{-1})^*$. The prior is again $\tilde{\varphi}_{m+1|m} \approx 0$. In the moist case, the posterior state that minimizes the quadratic functional is

$$\tilde{\varphi}_{m+1|m+1} \approx \mathbf{R}\mathbf{H}^*(\mathbf{H}\mathbf{R}\mathbf{H}^* + \mathbf{R}^o)^{-1}\hat{\psi}_{m+1} \quad (4)$$

In this case, with both $\mathbf{R}$ and $\mathbf{R}^o$ diagonal (and with $\mathbf{H} = \mathbf{I}$), the equations once again decouple, and the RMSE for the j th variable (identical to the posterior covariance in this perfect-model set-up) given by

$$RMSE_j = \left[\frac{r_j r_j^o}{r_j + r_j^o} \right]^{1/2}. \quad (5)$$

Dividing by r_j and using the fact that $r_j^o = \alpha\gamma r_j$ (with $\gamma = 1$ for $j = 2, 3, 4$) gives

$$\frac{RMSE_j}{r_j} = \left[\frac{\alpha\gamma}{1 + \alpha\gamma} \right]^{1/2}. \quad (6)$$

For the dry case, with non-diagonal background error covariance matrix $\tilde{\mathbf{B}}$, the posterior estimate is now given by

$$\tilde{\varphi}_{m+1|m+1} = \tilde{\mathbf{B}}\mathbf{H}^*(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^* + \mathbf{R}^o)^{-1}\hat{\psi}_{m+1}. \quad (7)$$

Denoting $\tilde{\mathbf{K}} = \tilde{\mathbf{B}}\mathbf{H}^*(\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^* + \mathbf{R}^o)^{-1}$, the MSE may be calculated using (7) along with $\hat{\psi}_{m+1} = \mathbf{H}\hat{\varphi}_{m+1} + \sigma_{m+1}^o$, and the fact that the observational noise is independent of the truth signal $\hat{\varphi}_{m+1}$:

$$\begin{aligned}
 E \left[(\tilde{\varphi}_{m+1|m+1} - \hat{\varphi}_{m+1})(\tilde{\varphi}_{m+1|m+1} - \hat{\varphi}_{m+1})^* \right] \\
 &= E \left[\tilde{\varphi}_{m+1|m+1} \tilde{\varphi}_{m+1|m+1}^* \right] - E \left[\tilde{\varphi}_{m+1|m+1} \hat{\varphi}_{m+1}^* \right] - E \left[\hat{\varphi}_{m+1} \tilde{\varphi}_{m+1|m+1}^* \right] + E \left[\hat{\varphi}_{m+1} \hat{\varphi}_{m+1}^* \right] \\
 &= \tilde{\mathbf{K}} E \left[\hat{\psi}_{m+1}, \hat{\psi}_{m+1}^* \right] \tilde{\mathbf{K}}^* - \tilde{\mathbf{K}} E \left[\hat{\psi}_{m+1}, \hat{\varphi}_{m+1}^* \right] - E \left[\hat{\varphi}_{m+1}, \hat{\psi}_{m+1}^* \right] \tilde{\mathbf{K}}^* + E \left[\hat{\varphi}_{m+1}, \hat{\varphi}_{m+1}^* \right] \\
 &= \left[\tilde{\mathbf{K}}(\mathbf{H}\mathbf{B}\mathbf{H}^* + \mathbf{R}^o) - \mathbf{B}\mathbf{H}^* \right] \tilde{\mathbf{K}}^* + \mathbf{B} - \tilde{\mathbf{K}}\mathbf{H}\mathbf{B}
 \end{aligned} \tag{8}$$

as shown in equation (56) in the main text.

As $\tilde{\mathbf{B}}$ is non-diagonal, the MSE from (8) for the j th variable is more complicated than in the moist case; we write it out explicitly only for the $j = 1$ variable, $\tilde{\mathcal{M}}$, since this variable is most impacted by the alternate background error covariance matrix. As $\tilde{\mathbf{B}}$ depends on $\tilde{Q}$, f , R_1 , κ , etc. through $\mathbf{T}_{kl}$ and $\tilde{\mathbf{T}}_{kl}$, all of these parameters may be expected to appear in the error; the end result for the RMSE is

$$\frac{\widetilde{RMSE}_1}{r_1} = \sqrt{\frac{\alpha\gamma}{(1+\alpha\gamma)} + \frac{\alpha^2\gamma^2 B_0 [(1+\alpha)(1+\alpha\gamma) + \alpha^2 B_0]}{(1+\alpha\gamma)[(1+\alpha)(1+\alpha\gamma) + \alpha B_0]^2}}, \tag{9}$$

where

$$B_0 = \tilde{Q}^2 \left[\frac{1}{f^2 R_1^2} \frac{r_2}{r_1} \right] \left[\frac{1 + (R_1 \kappa)^2 (r_3/r_2)}{1 + (R_1 \kappa)^2} \right]. \tag{10}$$

Thus the RMSE difference is

$$\frac{\widetilde{RMSE}_1 - RMSE_1}{RMSE_1} = \sqrt{1 + \frac{\alpha\gamma B_0 [(1+\alpha)(1+\alpha\gamma) + \alpha^2 B_0]}{[(1+\alpha)(1+\alpha\gamma) + \alpha B_0]^2}} - 1. \tag{11}$$