Data assimilation in the laboratory using a rotating annulus experiment

R. M. B. Young and P. L. Read

Supporting Information

Contents

1	Full MORALS equations	1
2	The analysis correction method	2
3	ACACIA — technical details	3
4	Observations - some technical points 4.1 ΔT in each datafile 4.2 Data pre-processing 4.3 Estimate of observational error 4.4 Weighting for verification	10 10 10 12 13
5	Optimizing the assimilation algorithm5.1 Regular flow (datafile expf2)5.2 Chaotic flow (datafiles expf5 and expf6)	14 14 16
6	Analysis increments for various assimilations	17
7	Residual error dependence on Ω — significance test	18

1 Full MORALS equations

The equations solved are the continuity, heat, and Navier-Stokes equations for an incompressible fluid:

$$\nabla \cdot \mathbf{u} = 0 \tag{1}$$

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = \nabla \cdot (\kappa \nabla T) \tag{2}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + 2\mathbf{\Omega} \times \mathbf{u} - (\Omega^2 R \hat{\mathbf{R}} - g \hat{\mathbf{z}}) \frac{\Delta \rho}{\rho_0} + \nabla \Pi = \mathbf{F}$$
(3)

T is relative to $T_R = 22 \text{ °C}$, and Π is relative to a reference pressure $\Pi_0(R, z) = \frac{1}{2}\Omega^2 R^2 + g(d-z)$. Here **F** is the viscous term, but is more complicated than the standard $\nu \nabla^2 \mathbf{u}$ because it allows spatial variations of the viscosity as well; see Farnell and Plumb (1976, Eqs 2.1–2.3, 2.6) for the full expression. These equations are closed by an equation of state for density

$$\rho(T) = \rho_0 (1 + \rho_1 T + \rho_2 T^2) \tag{4}$$

such that $\Delta \rho = \rho - \rho_0$, and two 'constitutive' relations for kinematic viscosity

$$\nu(T) = \nu_0 (1 + \nu_1 T + \nu_2 T^2) \tag{5}$$

and thermal diffusivity

$$\kappa(T) = \kappa_0 (1 + \kappa_1 T + \kappa_2 T^2) \tag{6}$$

The values used are

$\rho_0 =$	$1.043{ m gcm^{-3}}$	$\nu_0 =$	$1.620 \times 10^{-2} \mathrm{cm}^2 \mathrm{s}^{-1}$	$\kappa_0 =$	$1.290 \times 10^{-3} \mathrm{cm}^2 \mathrm{s}^{-1}$
$\rho_1 =$	$-3.070 imes 10^{-4} \mathrm{degC^{-1}}$	$\nu_1 =$	$-2.790 imes 10^{-2} \mathrm{degC^{-1}}$	$\kappa_1 =$	$2.330 \times 10^{-3} \mathrm{degC^{-1}}$
$\rho_2 =$	$-7.830 \times 10^{-6} \mathrm{degC^{-2}}$	$\nu_2 =$	$6.730 \times 10^{-4} \mathrm{degC^{-2}}$	$\kappa_2 =$	$0.000 \times 10^{-3} \mathrm{degC^{-2}}$

On all boundaries $\mathbf{u} = \mathbf{0}$ and $\hat{\mathbf{n}} \cdot \nabla \Pi = 0$, on the top and bottom boundaries $\hat{\mathbf{n}} \cdot \nabla T = 0$ (insulating), and on the inner and outer cylinders $T = T_a - T_R$ and $T = T_b - T_R$ respectively (conducting).

The model has 184,320 grid points $(24 \times 64 \times 24 \times 5)$, but the dimensionality of the model state space (the number of independent variables) is only 121,088. This is because (1) the Π field is diagnostic, and (2) there are a number of grid points on and outside the boundaries whose values are fixed by the boundary conditions. The Arakawa-C grid includes some points outside the boundary of the fluid.

2 The analysis correction method

We give a general outline of the analysis correction (AC) method below; full details can be found in Lorenc *et al.* (1991, hereafter L91). We use the standardised notation for data assimilation compiled by Ide *et al.* (1997) throughout, including where it conflicts with L91. (We made the following changes to the L91 notation: $\mathbf{c} \to \mathbf{d}, K \to H, \mathbf{K} \to \mathbf{H}, \mathbf{O} \to \mathbf{R}$, and $\Delta \mathbf{x} \to \delta \mathbf{x}$.)

The optimal analysis \mathbf{x}^{a} is a vector representing the best estimate of the true state of the system at time t^{a} . In MORALS the full system state, and hence the analysis, has dimension $5N_{R}N_{\phi}N_{z}$. This consists of five model fields \mathbf{u}^{a} , \mathbf{v}^{a} , \mathbf{w}^{a} , \mathbf{T}^{a} , and $\mathbf{\Pi}^{a}$, each with dimension $N_{x} = N_{R}N_{\phi}N_{z}$, where N_{R} , N_{ϕ} , and N_{z} are the number of model grid points in R, ϕ , and z.

In AC, as in most other formulations of data assimilation, the analysis minimises a cost function based on the errors in the different sources of information contributing to it. In AC the cost function is a trade-off in an RMS sense between the background state \mathbf{x}^{b} [length N_{x}] and the observations \mathbf{y}^{o} [N_{y}]. The optimal analysis minimises

$$J = \left[\mathbf{y}^{\mathrm{o}} - H(\mathbf{x})\right]^{*} \left(\mathbf{R} + \mathbf{F}\right)^{-1} \left[\mathbf{y}^{\mathrm{o}} - H(\mathbf{x})\right] + \left(\mathbf{x}^{\mathrm{b}} - \mathbf{x}\right)^{*} \mathbf{B}^{-1} \left(\mathbf{x}^{\mathrm{b}} - \mathbf{x}\right)$$
(7)

with respect to \mathbf{x} . \mathbf{B} is the background error covariance $[N_x \times N_x]$, \mathbf{R} is the observational error covariance $[N_y \times N_y]$, H is an interpolation operator mapping the model state \mathbf{x} to the observation points, and \mathbf{F} is the error covariance associated with H $[N_y \times N_y]$. This equation is solved in an iterative sense using a method of steepest descent, and making a number of approximations (L91, p.65) we obtain the successive corrections algorithm:

$$\mathbf{x}[n+1] = \mathbf{x}[n] + \mathbf{QW} \{\mathbf{y}^{o} - H(\mathbf{x}[n])\}$$
(8)

$$\mathbf{W} = \mathbf{B}\mathbf{H}^* \left(\mathbf{R} + \mathbf{F}\right)^{-1} \tag{9}$$

where $\mathbf{W} [N_x \times N_y]$ is a matrix of weights depending on the distance between observations and grid points, \mathbf{Q} is a matrix of normalization factors $[N_x \times N_x]$, and \mathbf{H} is a matrix of partial derivatives of H. Note that, in general, H is a nonlinear function, and that \mathbf{H} is a function of \mathbf{x}^{b} .

This method takes a first guess, compares it with the observed values at the observation points, and maps the difference between the two onto the model grid using various weighting functions. The observations used in each assimilation are taken from a window stretching forward in time by $t^{\rm b}$ and backwards in time by $t^{\rm f}$ from the analysis time, and each observation's contribution is weighted depending on how far the observation is from each grid point, and how close it is in time to the analysis time. Equation (8) does not distinguish in its weighting between observations in different directions, however, but only by distance from the grid point. Analysis correction improves this by replacing \mathbf{Q} with a matrix $\tilde{\mathbf{Q}}$ evaluated at the observation points $[N_y \times N_y]$, which accounts for the local observation density. The modified form, which the Met Office AC scheme used (L91, Eq. 3.9), is

$$\mathbf{x}[n+1] = \mathbf{x}[n] + \mathbf{W}\tilde{\mathbf{Q}} \left\{ \mathbf{y}^{\circ} - H(\mathbf{x}[n]) \right\}$$
(10)

$$\tilde{\mathbf{Q}} = (\mathbf{H}\mathbf{W} + \mathbf{I})^{-1} \tag{11}$$

Each run of ACACIA computes a single iteration step, so the form we use is

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{W}\tilde{\mathbf{Q}}\left\{\mathbf{y}^{\mathrm{o}} - H(\mathbf{x}^{\mathrm{b}})\right\}$$
(12)

The process of updating and repeated insertion nudges the analysis towards the optimal analysis. L91 (pp. 84–85) derive the gain matrix $\mathbf{W}\tilde{\mathbf{Q}}$ [$N_x \times N_y$] by combining Eqs (9) and (11):

$$\mathbf{W}\tilde{\mathbf{Q}} = \mathbf{B}\mathbf{H}^* \left(\mathbf{H}\mathbf{B}\mathbf{H}^* + \mathbf{R} + \mathbf{F}\right)^{-1}$$
(13)

which can be approximated in component form by (L91, Eqs A1.9 and 3.21)

$$[\mathbf{W}\tilde{\mathbf{Q}}]_{ki} = \frac{\mu_{ki}}{\varepsilon_i^2 + (1 + \varepsilon_i^2)^{1/2} D_i} \equiv \mu_{ki} \tilde{Q}_i$$
(14)

$$D_i \equiv H_i \left[\sum_j \mu_j R_j(\delta t_j) (1 + \varepsilon_j^2)^{-1/2} \right]$$
(15)

by the use of various approximations (L91, pp. 84) and a diagonalising approximation to the inverse term (Bratseth, 1986, Eqs 15a, 16, 18).

In this expression, μ_{ki} represents a spatial weighting function between model grid point k and observation i, D_i is the time-dependent data density at observation i (the term in square brackets is the equivalent at the model grid points, L91, Eq. 3.21), μ_j represents the weighting at all grid points affected by observation j, $R_j(\delta t_j)$ is the temporal weighting function between observation and analysis time, H_i interpolates from the model grid to observation i, and

$$\varepsilon_i^2 = \frac{\sigma_{ii}^2 + f_{ii}^2}{b_{ii}^2} \tag{16}$$

is the ratio of error variances for observation i, where σ_{ii}^2 is the observational error covariance (from **R**), f_{ii}^2 is the interpolation error covariance (from **F**), and b_{ii}^2 is the background error covariance interpolated to observation i (from **B**). AC approximates each of these error covariance matrices by its diagonal, which simplifies the expressions considerably. Note that the expression for D_i is slightly different from the expression derived by L91 in their Appendix 1. We use this form here because the sum is over all the observations in the assimilation window, and hence time weighting is necessary as otherwise all the observations are treated as if they are valid at the analysis time.

3 ACACIA — technical details

The aim in ACACIA is to use a method as close as possible to AC as described by L91. Figure S1 shows a flowchart of the whole procedure. The following sections describe each component of AC as applied to the rotating annulus case using ACACIA and MORALS.

Increments to the horizontal velocity field

Only horizontal velocity fields **u** and **v** are available as observations, so first ACACIA calculates the increments to these fields, and then dynamically balanced increments are added to the other fields (see below). Equation (12) is expressed in incremental form for grid point k by (L91, Eq. 3.18)

$$\delta x_k = \lambda \sum_i \mu_{ki} \tilde{Q}_i R_i^2(\delta t_i) d_i \tag{17}$$

where the first two terms inside the sum come from Eq. (14). $R_i^2(\delta t_i)$ represents the weighting of each observation in time (see below), and d_i is the increment at observation *i*:

$$d_i = y_i - H_i(\mathbf{x}^{\mathrm{b}}) \tag{18}$$



Figure S1: Flowchart showing the algorithm for computing a series of analyses with ACACIA.

For grid point k, the sum in Eq. (17) is over all observations i within the horizontal and vertical domains defined below in the section on weighting functions. δx_k represents either radial velocity δu_k or azimuthal velocity δv_k , which are calculated separately. These increments define the horizontal velocity analysis $\mathbf{u}^{\mathbf{a}} = \mathbf{u}^{\mathbf{b}} + \delta \mathbf{u}$ and $\mathbf{v}^{\mathbf{a}} = \mathbf{v}^{\mathbf{b}} + \delta \mathbf{v}$.

Relaxation coefficient λ

The term outside the sum in Eq. (17) is the *relaxation coefficient* λ , a scalar setting the fraction of the calculated increment to add to the background fields \mathbf{u}^{b} and \mathbf{v}^{b} . It is related to the nudging coefficient G (Hoke and Anthes, 1976) by

$$\lambda = \frac{G\Delta t}{1 + G\Delta t} \tag{19}$$

and performs a similar function, nudging the state towards the optimal analysis over a sequence of assimilations. Δt is the time between assimilations, and we set G to 4.45 times the Coriolis parameter. The relaxation coefficient and nudging implicitly perform a similar rôle to *initialisation* in other assimilation techniques.

We set G as follows: L91 state that the timescale G^{-1} should not be small compared with f^{-1} , where f is the Coriolis parameter. Let $\epsilon = |f/G|$ and suppose this ratio is the same for the atmosphere and the annulus.

In the Earth's atmosphere $f = 1.45 \times 10^{-4} \sin \phi$, where ϕ is latitude. Using the values of G in Appendix 2 of L91, we calculate ϵ for the Earth in three different cases: for the northern hemisphere $\epsilon \sim 0.171 \ (\phi \sim 45^{\circ})$, for the tropics $\epsilon \sim 0.081 \ (\phi \sim 10^{\circ})$, and for the southern hemisphere $\epsilon \sim 0.278 \ (\phi \sim -45^{\circ})$. The average is $\epsilon \sim 0.225 \ {\rm or} \ 1/\epsilon = 4.45$, so G = 4.45f.

Using $f = 2\Omega$, for the values of Ω and Δt we use, λ varies between 0.94 and 0.96. *G* is substantially larger in the annulus than the equivalent value for planetary atmospheres $(10^{-4}-10^{-3} \text{ s}^{-1})$ because the rotation rate is four orders of magnitude larger in the laboratory.

Background \mathbf{x}^{b} and its error covariance B

The background state \mathbf{x}^{b} represents the *a priori* estimate of the system state, and the background error covariance **B** is the estimate of its error. A MORALS simulation generates the background \mathbf{x}^{b} . For the first assimilation in a sequence, we ran MORALS for 1850 s, long enough for a coherent wave structure to appear and for any transient behaviour to decay, and used the final state in this run as the background state. For subsequent assimilations the background was the end point of a short range MORALS run started from the previous analysis.

In AC **B** is approximated by a diagonal matrix b_{ii}^2 , where b_{ii}^2 is the error variance at observation point *i*, interpolated from the model grid. We estimated the error on the model grid using a simple measure of variability in the MORALS background field at each radius and height. **B** was not updated between assimilations, as in general the observational error is much smaller. For radial velocity

$$B_u(R,z) = \frac{1}{N_\phi N_t} \sum_{\phi,t} \left[\mathbf{u}^{\mathrm{b}}(R,\phi,z,t) - \langle \mathbf{u}^{\mathrm{b}}(R,z) \rangle \right]^2$$
(20)

and similarly for \mathbf{v} , where N_t is the number of data points in time in the background run. In a sense this is a measure of 'climate' variability, although the amount of data used is much shorter than would normally be used to calculate a climate distribution. The rotational symmetry of the annulus means this quantity is a function of R and z only.

After the assimilations were completed another method for estimating **B** was suggested to us. After the first assimilation step the background is closer to the truth than the initial background state was, so except in the first assimilation **B** will be an overestimate. Instead of using a fixed covariance based on a free-running simulation, the alternative method would use the one-step forecast error (the difference between the background state and the assimilated observations) as an estimate of **B**. To confirm whether using this method would mean **B** had more of an effect on the assimilation, we calculated the ratio of one-step forecast error variance to observational error variance for each analysis. We found the original method that we did use suitable in almost all cases, as this ratio varied between 2:1 and 70:1 for regular flow and between 10:1 and 40:1 for chaotic flow, i.e. background error was always larger than observational error. The ratio was less than 5:1 only at a few times in the upper level radial velocity of the regular flow assimilations. In these cases **B** would have had an appreciable effect on the analysis increment and so the one-step error would have been a better estimate of that quantity.

Observations \mathbf{y}^{o} and their error covariance \mathbf{R}

The observations to be used for each assimilation are represented by a vector \mathbf{y}^{o} of length N_{y} . \mathbf{y}^{o} represents either radial velocity \mathbf{u}^{o} or azimuthal velocity \mathbf{v}^{o} , and contains observations for the whole assimilation window $(t^{a} - t^{f}$ to $t^{a} + t^{b})$.

Like **B**, AC assumes **R** $[N_y \times N_y]$ is diagonal. Section 4.3 below contains an estimate of the error from first principles, and concludes that the error distribution for each observation is approximately the same, with the following error covariance:

$$[\mathbf{R}]_{ii} \equiv \sigma_{ii}^2 = (0.0057 \,\mathrm{cm}\,\mathrm{s}^{-1})^2 \tag{21}$$

Background field transformation

The initial background **u** and **v** fields from the free-running MORALS simulation are generally out of phase with the observations at t^{start} . So, before calculating the first analysis increment in the sequence, ACACIA transforms \mathbf{x}^{b} in an azimuthal direction to align its main baroclinic wave with the observations at that time.



Figure S2: Radial velocity as a function of azimuth for observations interpolated to the mid-R / mid-z line (thick), and the background field (thin). In this example the background field was transformed by +5.77 rad to match up with the observations.

To calculate the transform angle the radial velocity is plotted at mid-radius $(R_{1/2})$ and observation level $z_{obs}(t^a)$ as a function of azimuth for both background and observed values. To calculate these azimuthal sections, ACACIA first interpolates the background radial velocity to the observation level using linear interpolation (there is a radial velocity grid point at mid-radius). Second, it estimates the observed radial velocity field at each mid-radius azimuthal grid point by taking a weighted mean over nearby observations:

$$u_{R_{1/2},\phi,z_{\rm obs}}^{\rm o} \simeq \frac{\sum_{r_{ji} \le r_{\rm max}} u_i^{\rm o} / r_{ji}^2}{\sum_{r_{ji} \le r_{\rm max}} 1 / r_{ji}^2}$$
(22)

where r_{ji} is the horizontal distance between observation u_i^{o} and the point $(R_{1/2}, \phi, z_{obs})$. The sum is over all the observations within r_{max} , a fixed multiple of the maximum distance between any grid point and its nearest observation. We found the results were not sensitive to this multiple, and used 1.1.

Figure S2a shows two such azimuthal sections found using this method. A spatial Fourier transform (Press *et al.*, 1992, p. 501) then extracts the phase of the dominant wave in both sections. Suppose the phase difference is $\Delta \theta_t = \theta^{\mathbf{u}^\circ} - \theta^{\mathbf{u}^\flat}$, where $\theta^{\mathbf{u}^\circ}$ and $\theta^{\mathbf{u}^\flat}$ are the phases in the observed and background fields respectively. ACACIA then transforms all five background fields by $+\Delta \theta_t$ to line up with the observations (Fig. S2b). The error associated with this transform is small compared with **B**, and **B** itself is not transformed because it is azimuthally symmetric.

Interpolation operator H and its error covariance F

ACACIA uses the interpolation operator H in a number of places to map a field from the model grid to the observation points. As the error in the interpolation is generally smaller than other sources of error, a simple 3D linear interpolation is sufficient. Cheney and Kincaid (2003, pp. 172–174) show that the error in a value obtained by linear interpolation has standard deviation

$$h = \frac{1}{8}(x_b - x_a)^2 \max_{x \in [x_a, x_b]} g''(x)$$
(23)

where g(x) is the function to be approximated by interpolation, and $[x_a, x_b]$ is the interval over which the interpolation is done. Assuming the second derivative at $(x_a + x_b)/2$ is representative of the whole interval, we can estimate the error in the interpolation over $[x_a, x_b]$ as

$$h_{[x_a,x_b]} \simeq \frac{x_b - x_a}{8} \left[\frac{g(x_{b+1}) - g(x_a)}{x_{b+1} - x_a} - \frac{g(x_b) - g(x_{a-1})}{x_b - x_{a-1}} \right]$$
(24)

Applying this to the linear interpolation operator H and combining errors in quadrature (Squires, 2001, p. 29) in each step of the linear interpolation in three dimensions, we obtain the interpolation error covariance \mathbf{F}_{ii} at each observation point i.

Weighting functions

The influence the *i*th observation has on the analysis at a particular model grid point depends on (1) the time between the observation time t_i^{o} and the analysis time t^{a} , and (2) the distance between the observation and the grid point.

For analysis time t^{a} , observation time t^{o}_{i} , window length forwards in time t^{f} and backwards in time t^{b} relative to each observation, the weighting for observation i is

$$R_i(\delta t_i) = \begin{cases} 1 - \delta t_i/t^{\rm b} & 0 \le \delta t_i \le t^{\rm b} \\ 1 + \delta t_i/t^{\rm f} & -t^{\rm f} \le \delta t_i < 0 \\ 0 & \text{otherwise} \end{cases}$$
(25)

where $\delta t_i = t_i^{\rm o} - t^{\rm a}$. From the point of view of the assimilation, the observations used are in a window running from $t^{\rm a} - t^{\rm f}$ to $t^{\rm a} + t^{\rm b}$. The weighting in space due to observation *i* at grid point *k* is

$$\mu_{ki} = W_{ki}^{\rm h} W_{ki}^{\rm v} \tag{26}$$

which is the product of horizontal and vertical weighting functions. We use the weighting formula from L91:

$$W_{ki}^{\rm h} = \left(1 + \frac{r_{ki}}{s^{\rm h}}\right) \exp\left(\frac{-r_{ki}}{s^{\rm h}}\right) \tag{27}$$

and similarly for the vertical, W_{ki}^{v} . This function is shown in Fig. S3. r_{ki} is the horizontal distance between grid point k and observation i, and $s^{\rm h}$ is the horizontal correlation scale. Equation (27) is unbounded, so a maximum radius of influence is imposed to save computational expense at $r_{ki} = \alpha s^{\rm h}$. The correlation scale defines the shape of the weighting function and its area of influence, and is optimized empirically (Sect. 5). For each observation its value depends on the time difference between observation and analysis. For maximum and minimum correlation scales $s_{\rm max}^{\rm h}$ and $s_{\rm min}^{\rm h}$ (horizontal) and $s_{\rm max}^{\rm v}$ and $s_{\rm min}^{\rm v}$ (vertical), the horizontal correlation scale is

$$s^{\rm h}(\delta t_i) = \begin{cases} s^{\rm h}_{\rm min} + \delta t_i (s^{\rm h}_{\rm max} - s^{\rm h}_{\rm min})/t^{\rm b} & 0 \le \delta t_i \le t^{\rm b} \\ s^{\rm h}_{\rm min} - \delta t_i (s^{\rm h}_{\rm max} - s^{\rm h}_{\rm min})/t^{\rm f} & -t^{\rm f} \le \delta t_i < 0 \\ N/A & \text{otherwise} \end{cases}$$
(28)

and similarly for the vertical.



Figure S3: Spatial weighting using Eq. (27) as a function of observation to grid point distance with a correlation scale of one and $\alpha = 3.5$.

Maximum influence radius α

By setting $\alpha < \infty$ in the sum over observations when calculating the u and v analysis increments, we omit some fraction of the total 'influence' of the observations on the analysis increment, because some observations are left out that would have had nonzero weight. We can estimate the fraction of influence omitted as a function of α assuming constant s and observation density. By integrating Eq. (27) over the area enclosed by $r \leq \alpha s$ (I_{α}) and comparing it with the integral over the area enclosed by $r \leq \infty$ (I_{∞}) we obtain the fraction of influence lost:

$$L_{\alpha} = \frac{I_{\infty} - I_{\alpha}}{I_{\infty}} = \left(1 + \alpha + \frac{\alpha^2}{3}\right) e^{-\alpha}$$
⁽²⁹⁾

The value eventually chosen for the assimilations, $\alpha = 5.92$, represents a 5% loss.

Balanced increments for the w, T, and Π fields

Increments to the other three model fields cannot be calculated directly from observational data, because no such data are available. ACACIA must instead use relationships between the model variables to calculate these increments and ensure consistency with the model dynamics.

In the interior of the fluid, the dominant balances are geostrophic (in the horizontal) and hydrostatic (in the vertical):

$$\frac{\partial \Pi}{\partial R} \approx fv \qquad \frac{\partial \Pi}{\partial \phi} \approx -fRu \qquad \frac{\partial \Pi}{\partial z} \approx -g\frac{\Delta \rho}{\rho_0} \tag{30}$$

We decided not to calculate balanced increments in temperature and pressure within the boundary layers, because the dominant balance there is significantly more complicated due to the inclusion of diffusion terms. ACACIA instead sets the temperature and pressure increments to zero within these layers, which extend about 1.25 mm into the fluid for typical values of ΔT and Ω . The Ekman layer thickness used to set the cutoff point for the top and bottom boundary layers is $BL_{Ek} = d\sqrt{Ek}$, where $Ek = \nu_0/(\Omega d^2)$ is the Ekman number, and the Stewartson / Pohlhausen layer thickness used to set the cutoff point for the inner and outer cylinder boundary layers is $BL_S = \min\left[(b-a)Ek^{1/3}, d/S^{1/4}\right]$, where $S = |\rho_1g(T_b - T_a)| d^3/(\nu_0\kappa_0)$ is the Rayleigh number.

Using the geostrophic balance relations for u and v and expanding in small increments, we obtain a Poisson equation for the pressure balance increment $\delta \Pi$ at each vertical level in the model:

$$\left[\frac{1}{R}\frac{\partial}{\partial R}\left(R\frac{\partial}{\partial R}\right) + \frac{1}{R^2}\frac{\partial^2}{\partial\phi^2}\right]\delta\mathbf{\Pi} = 2\Omega\left[\frac{\delta\mathbf{v}}{R} + \frac{\partial\delta\mathbf{v}}{\partial R} - \frac{1}{R}\frac{\partial\delta\mathbf{u}}{\partial\phi}\right]$$
(31)

This equation is $\mathbf{A} \delta \mathbf{\Pi} = \mathbf{b}$ in matrix form, and ACACIA solves this using the method of biconjugate gradients with \mathbf{A}^{\top} as a preconditioner matrix on both sides (Press *et al.*, 1992, p. 79), giving the pressure increment $\delta \Pi$ and hence the pressure analysis $\Pi^{a} = \Pi^{b} + \delta \Pi$. The integration constant is set so that the solution in the interior of the fluid matches up with the increment in the boundary layer (i.e. zero). If a particular level does not converge to a solution, ACACIA uses a linear interpolation between the nearest levels that do converge.

Using the hydrostatic approximation and the pressure increment $\delta \Pi$ we then obtain a quadratic equation for the temperature increment $\delta \mathbf{T}$ and hence the temperature analysis $\mathbf{T}^{a} = \mathbf{T}^{b} + \delta \mathbf{T}$. The full solution of this equation introduces rounding errors by subtracting a very small number from a very large one in the term under the square root, so we use the following approximation for the temperature increment at each grid point k:

$$\delta T_k = -\frac{1}{g \left[\rho_1 + 2\rho_2 T_k^{\rm b}\right]} \left. \frac{\partial \delta \mathbf{\Pi}}{\partial z} \right|_k \tag{32}$$

This is used as long as

$$\left[\frac{\rho_1}{2\rho_2} + T_k^{\rm b}\right]^2 \gg \frac{1}{g\rho_2} \left.\frac{\partial\delta\Pi}{\partial z}\right|_k \tag{33}$$

which is satisfied almost everywhere, otherwise the full quadratic equation is used.

Appendix D of Young (2009) contains a full derivation of these balance conditions and the finite-difference form of the Poisson equation used to solve Eq. (31).

We attempted to calculate a balanced vertical velocity increment $\delta \mathbf{w}$ using the continuity equation and the heat transfer equation, but both of these approaches proved inadequate. First, the continuity equation must be integrated across the fluid while satisfying zero velocity on both boundaries, which is not possible in practice. Second, the heat equation formulation includes a term $\left[\partial \mathbf{T}^{\mathrm{b}}/\partial z\right]^{-1}$, which is sometimes small, making $\delta \mathbf{w}$ sensitive to noise in $\delta \mathbf{u}$ and $\delta \mathbf{v}$. Instead, we set $\delta \mathbf{w}$ to zero everywhere, so $\mathbf{w}^{\mathrm{a}} = \mathbf{w}^{\mathrm{b}}$. This means the vertical velocity is updated only by the model, but this did not seem to cause any problems.

Algorithm procedure

The procedure for computing the analysis \mathbf{x}^{a} using ACACIA is as follows:

Read in command line parameters. • Read in parameters from a file. • Define the relaxation coefficient λ . • Read in the background field \mathbf{x}^{b} and its error covariance **B**. • Calculate the position of the boundary layers. • Read in velocity observations \mathbf{u}^{o} and \mathbf{v}^{o} within the assimilation window $[t^{a} - t^{f}, t^{a} + t^{b}]$, and convert them to plane polar coordinates. • Calculate the observational error covariance matrix R. • Calculate the weighting of each observation in time $R_i(\delta t_i)$. • If required, transform the background field to line up with the observations. • Plot the background fields and observations. • Interpolate the background error covariance **B** to the observation points. • Save the parameters, grids, background, and observations to disk. • Calculate the correlation scales for each observation. • Interpolate the background field to the observation points, $H(\mathbf{x}^{b})$. • Calculate the interpolation operator error covariance **F**. • Calculate the increments at the observation points, d. • Calculate the ratio of error variances ε_i^2 at the observation points. • Calculate the data density on the model grid using contributions from each observation calculated with the temporal and spatial weighting functions. • Interpolate the data density to the observation points to get D_i . • Calculate the approximation to \tilde{Q}_i using the data density D_i . • For each observation, add its contribution to the increments $\delta \mathbf{u}$ and $\delta \mathbf{v}$. • Calculate the balanced pressure increment $\delta \mathbf{\Pi}$. • Calculate the balanced temperature increment $\delta \mathbf{T}$. • Increment the velocities to get \mathbf{u}^{a} and \mathbf{v}^{a} and impose the boundary conditions. • Increment the pressure field to get $\Pi^{\rm a}$ and impose the boundary conditions. • Increment the temperature field to get $\mathbf{T}^{\mathbf{a}}$ and impose the boundary conditions. \bullet Plot the analysis increments, analysis fields, analysis divergence and vorticity fields, and vector velocity field of observations and analysis at the observation level.

• Output the analysis increments $\delta \mathbf{u}, \delta \mathbf{v}, \delta \mathbf{T}$, and $\delta \mathbf{\Pi}$ as a perturbation for MORALS.

Comparison of algorithm parameters with other AC uses

Table 1 lists typical parameter values used in previous applications of analysis correction and approximate equivalent values for the annulus in the regular flow assimilations.

In both these previous cases the model advances by just one time step between assimilations. For the annulus, however, this is prohibitively expensive because there are typically several hundred model time steps between each set of observations. Instead of running an assimilation every timestep ACACIA runs every 100–250 timesteps, or 2–5 times between observation datasets.

Parameter	Met Office	AOPP / OU	Annulus (equivalent)	Annulus (used)
N_x	10^{7}	10^{6}	121088	121088
N_y	100000	10000	9000	9000
t^{b}	$5\mathrm{h}$	$5\mathrm{h}$	$1.3\mathrm{s}$	$26\mathrm{s}$
t^{f}	$1\mathrm{h}$	$1\mathrm{h}$	$0.3\mathrm{s}$	$26\mathrm{s}$
$s_{\min}^{ m h}$	$300\mathrm{km}$	$340\mathrm{km}$	$0.35\mathrm{cm}$	$0.21\mathrm{cm}$
s_{\max}^{h}	$400\mathrm{km}$	$540\mathrm{km}$	$0.47\mathrm{cm}$	$0.42\mathrm{cm}$
s^{v}	3.0	3.0	N/A	$0.500.75\mathrm{cm}$
α	3.5	3.5	3.5	5.92
f	$10^{-4} \mathrm{rad} \mathrm{s}^{-1}$	$10^{-4} \mathrm{rad} \mathrm{s}^{-1}$	2Ω	2Ω
G	$4 \times 10^{-4} \mathrm{rad}\mathrm{s}^{-1}$	$5 \times 10^{-4} \mathrm{rad}\mathrm{s}^{-1}$	8.9Ω	8.9Ω
Δt	$15\mathrm{min}$	$7.5\mathrm{min}$	$0.065\mathrm{s}$	$2.5\mathrm{s}$
δt	$15\mathrm{min}$	$7.5\mathrm{min}$	$0.065\mathrm{s}$	$0.02\mathrm{s}$
λ	0.25	0.2	0.96	0.96

Table 1: Typical parameters for the Met Office operational (Lorenc *et al.*, 1991) and AOPP / OU Mars (Lewis *et al.*, 1996; Montabone *et al.*, 2006) analysis correction schemes, equivalent annulus values scaling down from the Met Office values, and the range of values eventually used in the annulus assimilations. N_y for Earth is from Kalnay (2003, p. 13). Note s^v is defined differently in the Met Office method.

4 Observations - some technical points

The observations may be available from the authors on request.

4.1 ΔT in each datafile

Figure S4 shows the temperature difference between the two cylinders as a function of time for each of the datafiles.

4.2 Data pre-processing

First we discarded the first 225 datasets from datafile expf5, because these datasets were separated by less than 5 s, and the algorithm is much simpler to code if the time between each dataset is the same.

Forcing parameters

For each dataset two raw measurements exist for Ω and ΔT . We pre-processed these measurements by using the mean of the two values unless they differed by more than 2% or differed from the value for the previous dataset by more than 2%. In these cases any spurious measurements (occasionally values of zero were recorded) were removed manually.



Figure S4: Temperature difference $\Delta T = T_b - T_a$ as functions of time in each of the experimental datafiles. Note that the scatter probably represents measurement error and not real scatter in the forcing.

Both ΔT and Ω used by each ACACIA step are extracted from the observational dataset for the observations valid at the analysis time (or the closest observations in time, if the analysis time does not coincide with any observations).

Velocities

The number of raw and processed velocity observations in each datafile are:

Datafile	expf2	expf5	expf6
Observations (raw)	1207546	3577702	2408183
Observations (processed)	1206296	2877609	2405305
Datasets (processed)	2150	1822	2200
Datasets at each vertical level	430	911	1100

To pre-process the velocity observations, we first removed any (erroneously recorded) observations outside the horizontal boundaries of the fluid. We then used an automated filter to identify the 15 nearest neighbours for each observation. The filter calculated the mean and standard deviation of their velocities, and removed the observation if either velocity component fell outside $\pm 5\sigma$ from the mean of the 15 neighbours. The filter did not process observations within 5% of the radius of either boundary, as velocities there are small and observational error has a larger percentage effect on the spread. This automated filter erroneously removed about 400–500 good observations, but this is less than 0.1% of the total, an acceptable loss of good data if all the bad observations are removed as well.

A final step was required after running the filter. Because the velocity and parameter observations were recorded independently they recorded different times for their measurements. This was resolved by manually identifying the Ω transitions in the velocity data and matching these to transitions in the measured Ω parameter, and then redefining the time of the first dataset to be zero.

Subsets and missing data

As noted in the main text, to simplify the coding we assumed the subsets in a particular dataset were valid at the same time. Because the zero point of the datafile's time measurement is arbitrary, we defined this time to be at the start of the 5 s window. The times corresponding to the various subsets are listed below.

A few datasets were missing from datafiles expf5 and expf6. The sequences of assimilations were chosen so as not to overlap with any of the missing datasets. The missing observational datasets for each datafile and subset were:

Datafile	Subset	Time	Missing datasets (times in s)
expf2	1 blue	$2.8\mathrm{s}$	None
	$2 \operatorname{red}$	$3.6\mathrm{s}$	None
expf5	1 blue	$0.7\mathrm{s}$	1295,1320,2230,2915,3495,6365,7850,8735,9055,9075,9085,9100,9105
	$2 \operatorname{red}$	$1.5\mathrm{s}$	$375,\ 960,\ 1295,\ 1320,\ 1905,\ 2210,\ 2230,\ 3215,\ 3495,\ 6365,\ 7800,\ 8735,\ 9055,\ 9085,\ 9095$
	3 orange	$2.3\mathrm{s}$	375,960,1320,1905,2210,2975,3215,4825,7825,7875,8230,8735,9055,9085
expf6	1 blue	$0.9\mathrm{s}$	1075,6535,9795,10885,10915
	$2 \operatorname{red}$	$1.9\mathrm{s}$	1075,5645,6795,9935,10695,10885,10895,10915

4.3 Estimate of observational error

We used a simple estimate of the observational error; in reality a full observational error analysis would be more complicated than presented here. For the sake of a little extra accuracy this is not worth the effort, however, because the observational error covariance \mathbf{R} is generally much smaller than the background error covariance \mathbf{B} and so the assimilation converges towards the observed values anyway.

Dig*Image* (Dalziel, 1995) records the observations as sets of images over a $N_x \times N_x$ pixel grid (Fig. S5a). A single particle's position is defined as the centre of the pixel, shown in Fig. S5b. Assuming a uniform distribution of particles, the position of a single particle is uniformly distributed within a range $2b/N_x$ in each direction, where b is the outer cylinder radius, so the error in a single particle's position has mean zero and standard deviation $\sigma_x^s = \sigma_y^s = b/(\sqrt{3}N_x)$ using the properties of the uniform distribution, and the error is isotropic. Each observation position (x_i^o, y_i^o) is a mean over n^s snapshots $(n^s = 5 \text{ here})$, which reduces the positional error by a factor of $\sqrt{n^s}$. So we obtain the positional error

$$\sigma_{x_i^{\rm o}} = \sigma_{y_i^{\rm o}} = \frac{b}{N_x \sqrt{3n^{\rm s}}} \tag{34}$$

Each component of the velocity is given by the distance the particle has moved over the observation period t^{obs} divided by that period. $t^{\text{obs}} = 1 \text{ s}$ and is assumed to be exact. The speed in each direction is given by the difference in position between the start and end of the particle track, divided by t^{obs} , so this gives an error of

$$\sqrt{\left(\frac{\sigma_x^{\rm s}}{t^{\rm obs}}\right)^2 + \left(\frac{\sigma_y^{\rm s}}{t^{\rm obs}}\right)^2} = \sqrt{2}\frac{\sigma_x^{\rm s}}{t^{\rm obs}} \tag{35}$$

for a single pair of start and end points. The velocity is averaged over a number of particle pairs in the track, approximately $n^{\rm s}$, which reduces the velocity error by a factor of $\sqrt{n^{\rm s}}$. The velocity error is isotropic



Figure S5: Schematic of the observation grid and a single pixel.

as well, so the error in (R, ϕ) velocity (the form required by the assimilation) is the same as in (x, y) velocity (in which the observations are recorded). Substituting in for σ_x^s we obtain the velocity error

$$\sigma_{u_i^{\rm o}} = \sigma_{v_i^{\rm o}} = \frac{b}{N_x t^{\rm obs}} \sqrt{\frac{2}{3n^{\rm s}}} \tag{36}$$

and hence the observational error covariance is a diagonal matrix

$$\mathbf{R} = \frac{2}{3n^{\mathrm{s}}} \left(\frac{b}{N_X t^{\mathrm{obs}}}\right)^2 \mathbf{I}$$
(37)

Putting in the values $N_x = 512$, $t^{\text{obs}} = 1$ s, $n^s = 5$, and b = 8.0 cm, this gives errors in the observed positions of 0.0040 cm and velocities of 0.0057 cm s⁻¹. Note that although the observation error is uniformly distributed the assimilation assumes a normal distribution of error, so the approximation is made that the standard deviation of the two is the same. In practice the error distribution is approximate enough for this to be OK. The assimilation also assumes that the positions of the observations are measured exactly, so the error in the observed positions doesn't enter into the calculations.

4.4 Weighting for verification

We verify the analyses against observations distributed irregularly in space. For verification scores representative of the whole domain (such as residual error), it is generally not appropriate to consider all observations equivalent, because they are clustered in some places and sparse in others, and an observation in a sparse region is representative of a larger area. To account for this we assigned the *n*-th analysis-verification pair a weight $w_n = 1/d_n$, where d_n is the local observation density. The total weight over N pairs is $W = \sum_{n=1}^{N} w_n$. For an observation at Cartesian coordinates (x_n, y_n) with $r_n = \sqrt{x_n^2 + y_n^2}$, we define the data density to be the number of observations within r_c of the observation, divided by the area enclosed by r_c :

$$d_n = \frac{1}{A} \sum_{x,y} \left[\left\{ (x - x_n)^2 + (y - y_n)^2 \right\} \le r_c^2 \right]$$
(38)

where the sum is over all observations at that time, and A is the area. r_c is a representative scale, and tests showed that the distribution of weight is not particularly sensitive to its value, so it was set to 1 cm. The

Table 2: List of parameters tested to find the optimal assimilation parameters for the regular flow regime assimilations. Starting the assimilation at different observation levels and changing the multiple of the maximum distance to the nearest neighbour to use in the background transform were also tested, but made no difference.

Parameter	Symbol	Values	Comments
Assimilation time / s	Δt	0.2, 0.5, 1.0, 2.5, 5.0	
Assimilation window length $/$ s	$t^{\mathrm{b}} + t^{\mathrm{f}}$	24, 48, 96, 192	(with $t^{\rm b}/t^{\rm f} = 3$)
Assimilation window asymmetry	$t^{ m b}/t^{ m f}$	1/3, 0.5, 1, 2, 3, 3.8, 5	(with $t^{\rm b} + t^{\rm f} = 48 \mathrm{s}$)
Horizontal correlation scales / cm	$s_{ m min}^{ m h}$	0.1, 0.21, 0.26, 0.65, 1.61	
	$s_{ m max}^{ m h}$	0.2, 0.42, 0.42, 2.6, 2.6	(as pairs)
Vertical correlation scales / cm	s_{\min}^{v}	0.5, 0.5, 1.6, 3.15	
	$s_{ m max}^{ m v}$	0.75, 3.15, 3.6, 3.6	(as pairs)
Maximum influence cutoff	α	2.33, 3.5, 5.92	

area depends on how close the point is to the inner or outer boundaries; elementary geometry gives us:

$$A = \begin{cases} r_c^2(\pi - \phi_1) + b^2 \phi_2 - br_n \sin \phi_2 & b - r_c < r_n \\ \pi r_c^2 & a + r_c \leq r_n \leq b - r_c \\ r_c^2(\pi - \phi_3 + \sin \phi_3 \cos \phi_3) - a^2(\phi_4 - \sin \phi_4 \cos \phi_4) & r_n < a + r_c \end{cases}$$
(39)

where

$$\cos\phi_1 = \frac{r_c^2 + r_n^2 - b^2}{-2r_c r_n} \quad \cos\phi_2 = \frac{b^2 + r_n^2 - r_c^2}{2br_n} \quad \cos\phi_3 = \frac{r_c^2 + r_n^2 - a^2}{2r_c r_n} \quad \cos\phi_4 = \frac{a^2 + r_n^2 - r_c^2}{2ar_n} \tag{40}$$

5 Optimizing the assimilation algorithm

Several parameters in the assimilation can be tuned to optimize its accuracy. AC is usually optimized in an empirical sense, where the 'best' parameter values are those which give the most accurate analysis. We describe below a range of tests to find the most accurate combination of parameters. The accuracy of the analysis was judged by (a) the residual error, (b) the number of assimilation cycles required for the observations and background to converge, indicated by the first minimum in the residual error curve, (c) the real time taken to compute each assimilation, and (d) a visual comparison of the analysis and observations, to note any systematic errors. Criterion (b) can also be evaluated by measuring the RMS size of the increment in each analysis, which also falls to a minimum on convergence. For radial velocity it is given by

$$u_{\rm inc} = \sqrt{\sum_{ijk} \delta u_{ijk}^2 \Delta V_{ijk}} / \sum_{ijk} \Delta V_{ijk}$$
(41)

where δu_{ijk} is the radial velocity increment at the analysis time at position (i, j, k) on the model grid. Weighting is by ΔV_{ijk} , the model grid volume element at (R_i, ϕ_j, z_k) , and v_{inc} is defined similarly.

5.1 Regular flow (datafile expf2)

We ran 50s tests starting at $t^{\rm start} = 8335 \,\mathrm{s}$ and finishing at $t^{\rm stop} = 8385 \,\mathrm{s}$ in expf2 (2S regime, $\Omega = 0.775 \,\mathrm{rad}\,\mathrm{s}^{-1}$), using subset blue for assimilation and subset red for verification. The following parameters were used as a 'standard', changing one at a time: assimilation time $\Delta t = 0.5 \,\mathrm{s}$, window length backwards in time $t^{\rm b} = 36 \,\mathrm{s}$, window length forwards in time $t^{\rm f} = 12 \,\mathrm{s}$, minimum and maximum horizontal correlation scales $s^{\rm h}_{\rm min} = 0.26 \,\mathrm{cm}$ and $s^{\rm h}_{\rm max} = 0.42 \,\mathrm{cm}$, minimum and maximum vertical correlation scales $s^{\rm v}_{\rm min} = 0.50 \,\mathrm{cm}$ and $s^{\rm v}_{\rm max} = 0.75 \,\mathrm{cm}$, and maximum influence cutoff $\alpha = 3.5$. We varied each of these parameters using the



Figure S6: Radial velocity analysis increments (Eq. 41) in the ACACIA 2S regime test varying assimilation time Δt . The lines correspond to 0.2s (solid), 0.5s (dotted), 1.0s (dashed), 2.5s (dot-dash), and 5.0s (dot-dot-dot-dot-dash).

values in Table 2, while keeping all the others constant. We made the assumption that each parameter can be varied independently before combining all the optimum values. This is required because it is not practical to test all possible combinations of parameter values.

Assimilation time Δt (time between assimilations): Overall, for u, longer assimilation times give better analyses (i.e. lower residual errors), and for v the opposite is the case. The u result was unexpected, and we have not yet found an explanation for it. This result is generally true at each vertical level, so overall the best choice is the intermediate 1.0 s. To cut down on the amount of data generated, however, 2.5 s was chosen. The convergence time depends approximately linearly on the assimilation time, so a similar number of assimilations are required to converge in each case (Fig. S6).

Assimilation window length t_b+t_f : In general (both u and v, at most levels), the shorter the window the smaller the residual error. The convergence rate is approximately the same in each case. Hence one should use the shortest possible window length (subject to a point made below about the window asymmetry). As the window length increases, data is used from further into the future and past relative to the analysis time. If the baroclinic wave has nonzero azimuthal phase speed, the window includes observations from times when the wave is out of phase with its position at $t = t^a$. This phase difference grows as the window length increases, and hence the accuracy decreases as the wave is 'smeared' out.

Assimilation window asymmetry t_b/t_f : At the different levels there were different results, but at each level there was a definite order with either the highest or lowest t^b/t^f giving the smallest residual error, although the differences were marginal. To compromise between these results a symmetrical window was chosen, which is quite different from L91, who used an asymmetric window with $t^b/t^f = 5$. Shorter assimilation windows give better results, but we also want some observations within the window at each level both forward and backward in time from t^a . The shortest window that guarantees this is $t^f = t^b = 26$ s.

Horizontal correlation scales s_{\min}^{h} and s_{\max}^{h} : These correlation scales should represent the general size of the features in the flow. The tests showed that the larger s^{h} is the longer the convergence time. In general small correlation scales were better, but there was not much to choose between these smaller values. For the small values, over all the vertical levels, comparing $[s_{\min}^{h}, s_{\max}^{h}] = [0.1 \text{ cm}, 0.2 \text{ cm}]$ with [0.21 cm, 0.42 cm] (same $s_{\max}^{h}/s_{\min}^{h}$ ratio) gave better results for higher s_{\min}^{h} . Comparing [0.21 cm, 0.42 cm] with [0.26 cm, 0.42 cm] (same s_{\max}^{h} , varying s_{\min}^{h}) gave similar results for both. Hence, $s_{\max}^{h}/s_{\min}^{h} = 2$ was chosen with $s_{\min}^{h} = 0.21$ and $s_{\max}^{h} = 0.42$, as the smaller s_{\min}^{h} required marginally less computing resources.

Parameter	Symbol	Values	Comments
Assimilation time / s	Δt	0.5, 1.0, 2.5	
Assimilation window length $/$ s	$t^{\mathrm{b}} + t^{\mathrm{f}}$	22, 26, 42, 62	(with $t^{\rm b} = t^{\rm f}$)
Assimilation window asymmetry	$t^{ m b}/t^{ m f}$	4/9, 1, 2.25, 3 1/3	(with $t^{\rm b} + t^{\rm f} = 26 {\rm s}$)
Horizontal correlation scales / cm	$s_{\min}^{ m h}$	0.10, 0.15, 0.21, 0.30	
	s_{\max}^{h}	0.20, 0.30, 0.42, 0.60	(as pairs)

Table 3: List of parameters tested to find the optimal assimilation parameters for the chaotic data.

Vertical correlation scales s_{\min}^{v} and s_{\max}^{v} : The vertical correlation scales depend on the vertical positions of the observations, because as this scale increases whole new datasets may be added to the observation vector. The scales must be large enough so that all the model levels are influenced by at least one set of observations. Over the range of values tested, larger vertical correlation scales gave longer convergence. For v the residual increases as s^{v} increases, but for u the order is different at different levels, so it was decided to use the smallest values $s_{\min}^{v} = 0.50 \text{ cm}$ and $s_{\max}^{v} = 0.75 \text{ cm}$. With the values of α used, even these small values ensured all model levels were influenced by at least one set of observations.

Maximum observation influence cutoff α : The tests used $\alpha = 2.33$ (50% loss of observational 'influence' – see earlier), 3.5 (26% loss), and 5.92 (5% loss). Considering all five observation levels together, smaller α gave a larger u residual error, and larger α gave a larger v residual, but the differences were negligible. The higher values gave smoother analysis fields, however, so $\alpha = 5.92$ was chosen.

Summary of test results: In most of the tests there was only a marginal difference between the poorest result and the best result, suggesting that most of the values scaled down from equivalent Earth values (Table 1) are generally good. The optimal parameter values chosen for the regular regime assimilations were as follows: $\Delta t = 2.5 \text{ s}$, $t^{\text{f}} = t^{\text{b}} = 26 \text{ s}$, $s^{\text{h}}_{\text{min}} = 0.21 \text{ cm}$, $s^{\text{h}}_{\text{max}} = 0.42 \text{ cm}$, $s^{\text{v}}_{\text{min}} = 0.50 \text{ cm}$, $s^{\text{v}}_{\text{min}} = 0.75 \text{ cm}$, and $\alpha = 5.92$.

5.2 Chaotic flow (datafiles expf5 and expf6)

Initially we ran the first assimilation in the chaotic regime using the optimal parameters determined in the previous section. This was assimilation m3sv1 and ran for 500s in the expf5 datafile at $\Omega = 2.3 \,\mathrm{rad \, s^{-1}}$ starting at $t = 2250 \,\mathrm{s}$ and finishing at $t = 2750 \,\mathrm{s}$. t^{f} and t^{b} were changed to 13s as there are only 10s between consecutive observations at each level in the chaotic data, and 13s is the smallest window that always includes at least one dataset at each level forward and backwards from the analysis time.

However, this assimilation produced analyses with each cyclonic lobe at the upper level split into two eddies (Fig. S7, left), while the observations suggest a single vortex. This problem is due to holes in the distribution of observations at three places in the flow (Fig. S7, right; similar holes exist at z = 4.3 cm), which are due to the neutrally buoyant particles falling out of suspension and forming a pile on the bottom of the tank. The lack of constraining observations in this region then led to a poor analysis, and as this cannot be fixed within the observational data it can only be fixed within the assimilation, by finding a set of parameter values that produces a better analysis in this regime.

We ran several 100 s assimilations to test the sensitivity of the analysis to the parameters again, starting at t = 2250 s in expf5 ($\Omega = 2.3$ rad s⁻¹), and using the red observational subset for assimilation and the orange subset for verification. Table 3 lists the values tested. The set of standard parameters were the same as for regular flow, but with $t^{\rm b} = t^{\rm f} = 13$ cm and increased vertical correlation scales $s_{\rm min}^{\rm v} = 0.75$ cm and $s_{\rm max}^{\rm v} = 1.00$ cm to ensure the observations affect all model grid levels (there are only two observation levels in the chaotic datafiles).

Assimilation time Δt (time between assimilations): A shorter assimilation time gave a better analysis for both u and v in terms of residual error. This result was more pronounced at z = 9.7 cm than at 4.3 cm.



Figure S7: Left: Detail of analysis m3sv1 at z = 9.7 cm, $t^{a} = 2660$ s. The westerly jet stream can be seen in the observations (red vectors) but there is an artefact of the assimilation in the analysis (blue vectors) at the centre of the cyclone. Right: Instantaneous velocity streamlines showing all the velocity vectors between t = 2500 s and t = 2700 s in datafile expf5 at z = 9.7 cm. Observations with zero velocity appear as dots. The three holes in the observations are clear.

The improvement from 2.5 s to 1.0 s was substantial (10%), and from 1.0 s to 0.5 s less so (2%), so we chose $\Delta t = 1.0$ s as a compromise between accuracy and speed.

Window length $t^{b} + t^{f}$: The results were different for each level and also between u and v. In the v velocity a shorter window led to a poorer assimilation, and in the u velocity at z = 9.7 cm a shorter window led to a better assimilation (negligible effect at z = 4.3 cm).

Window asymmetry t^b/t^f : In the *u* velocity the difference was marginal, but higher t^b/t^f was generally found to be better. In the *v* velocity the opposite effect was true, so we chose a symmetrical window again, with $t^b = 21$ s and $t^f = 21$ s. (There was only marginal improvement in making the window longer than this, but with a linear increase in runtime.) The reason for the difference in results between the *u* and *v* velocities is not clear.

Horizontal correlation scales s_{\min}^{h} and s_{\max}^{h} : In all cases smaller correlation scales led to better assimilations, so we chose $s_{\min}^{h} = 0.1 \text{ cm}$ and $s_{\max}^{h} = 0.2 \text{ cm}$.

Summary of test results: The revised parameters are $\Delta t = 1.0$ s, $t^{\rm b} = t^{\rm f} = 21.0$ s, $s^{\rm h}_{\rm min} = 0.1$ cm, $s^{\rm h}_{\rm max} = 0.2$ cm, $s^{\rm v}_{\rm min} = 0.75$ cm, $s^{\rm v}_{\rm max} = 1.00$ cm, and $\alpha = 5.92$.

6 Analysis increments for various assimilations

Figure $\frac{58}{8}$ shows RMS analysis increments (Eq. 41) from various assimilations described in the main text.

In m2s1 (Figs S8a, b) the v velocity increment is about double the u velocity increment, and the assimilation converges to an optimum analysis in less than 25 s. In the 3AV assimilations (not shown) the velocity increments are about the same size with approximately the same convergence time.

In m3av2s1 (Figs S8c, d) there is a large jump near the start of the assimilation. The rise begins at $t \sim 5725$ s, reaches a maximum at $t \sim 5750$ s, and subsides by $t \sim 5800$ s. The first two times correspond

to the first time observations from beyond the Ω transition enter the assimilation window (as $t^{\rm b} = 26 \, {\rm s}$), and the time of the Ω transition itself. The third time shows that it takes about 50 s for the assimilation to recover from a change in rotation rate. Note that this rotation rate transition has only a small effect on the residual error, much less than in assimilation m3sv3a.

Finally, in m3sv3a (Figs S8e, f) the rapid increase in the size of the analysis increment as Ω changes abruptly is clear.

7 Residual error dependence on Ω — significance test

Fig. S9 shows the residual error as a function of rotation rate for u and v, using all regular regime assimilations except m3av2s1. At first glance the plots seem to show some weak positive slopes, but because each point in each line has an error associated with it, a proper analysis requires the distribution of possible slopes. We can estimate the distribution of slopes using a bootstrap method. For each line, we generated a sample line using the means and standard deviations at each Ω . From this we calculated the slope of the line using the IDL routine LINFIT. This was repeated 500 000 times to give a distribution of possible slopes for each line, from which the 2.5% and 97.5% percentiles were extracted to give 95% confidence limits on the slopes of the lines. The result was that in no cases did 'zero slope' fall outside the 95% confidence intervals, so at this significance level there is no trend (positive or negative) between the residual error and the annulus rotation rate.

References

- Bratseth AM. 1986. Statistical interpolation by means of successive corrections. *Tellus* **38A**: 439–447, doi:10.1111/j. 1600-0870.1986.tb00476.x.
- Cheney W, Kincaid D. 2003. Numerical Mathematics and Computing. Thomson Brooks/Cole, 5th edn, ISBN 0534389937.
- Dalziel S. 1995. 'DigImage: Image processing for fluid dynamics Particle tracking'. URL http://www.damtp.cam. ac.uk/lab/digimage/. Version 2.0, 28 pp.
- Farnell L, Plumb RA. 1976. 'Numerical integration of flow in a rotating annulus II: three dimensional model'. Occasional Note Met O 21 76/1, Met Office: Exeter, UK. 24pp.
- Hoke JE, Anthes RA. 1976. The initialization of numerical models by a dynamic-initialization technique. Mon. Weather Rev. 104: 1551–1556, doi:10.1175/1520-0493(1976)104(1551:TIONMB)2.0.CO;2.
- Ide K, Courtier P, Ghil M, Lorenc AC. 1997. Unified notation for data assimilation: operational, sequential and variational. J. Meteorol. Soc. Jpn 75: 181–189.
- Kalnay E. 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, ISBN 0521796296. xxii+341pp.
- Lewis SR, Read PL, Collins M. 1996. Martian atmospheric data assimilation with a simplified general circulation model: orbiter and lander networks. *Planet. Space Sci.* 44: 1395–1409, doi:10.1016/S0032-0633(96)00058-X.
- Lorenc AC, Bell RS, Macpherson B. 1991. The Meteorological Office analysis correction data assimilation scheme. Q. J. R. Meteorol. Soc. 117: 59–89, doi:10.1002/qj.49711749704.
- Montabone L, Lewis SR, Read PL, Hinson DP. 2006. Validation of Martian meteorological data assimilation for MGS/TES using radio occultation measurements. *Icarus* 185: 113–132, doi:10.1016/j.icarus.2006.07.012.
- Press WH, Flannery BP, Teukolsky SA, Vetterling WT. 1992. Numerical Recipes in Fortran 77: The Art of Scientific Computing. Cambridge University Press, 2nd edn, ISBN 052143064X.
- Squires GL. 2001. Practical Physics. Cambridge University Press, 4th edn, ISBN 0521779405. xi+212pp.
- Young RMB. 2009. Predictability of a laboratory analogue for planetary atmospheres. DPhil thesis, University of Oxford.



Figure S8: RMS analysis increments averaged over the whole volume for regular assimilations m2s1 (2S, top), m3av2s1 (3AV to 2S transition, middle), and chaotic assimilation m3sv3a (3SV assimilation over a rotation rate transition, bottom).



Figure S9: u and v residual errors at each level as a function of rotation rate Ω for the regular regime assimilations. The mean residual error over all analyses is plotted as a dot, with the standard deviation as error bars. Results from assimilation m3av2s1 are not included.