

Predictability of the thermally driven laboratory rotating annulus

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

Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, UK

*Correspondence to: R. M. B. Young, AOPP, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK.
E-mail: roland.young@physics.ox.ac.uk

We investigate the predictability of the thermally driven rotating annulus, a laboratory experiment used to study the dynamics of planetary atmospheres under controlled and reproducible conditions. Our approach is to apply the same principles used to predict the atmosphere in operational weather forecasting. We build a forecasting system for the annulus using the analysis correction method for data assimilation, the breeding method for ensemble generation, and the Met Office/Oxford Rotating Annulus Laboratory Simulation as the forecast model. The system forecasts the annulus in steady (2S), amplitude vacillating (3AV), and structurally vacillating (3SV) flow regimes, verifying the forecasts against laboratory data. The results show that a range of flow regimes from this experiment can be accurately predicted. Forecasts in the steady wave flow regime perform well, and are predictable until the end of the available data. Forecasts in the amplitude and structural vacillation flow regimes lose quality and skill by a combination of wave drift and wavenumber transition. Amplitude vacillation is predictable up to several hundred seconds ahead, and structural vacillation is predictable for a few hundred seconds. The wavenumber transitions are partly explained by hysteresis in the rotating annulus experiment and model.

Key Words: rotating annulus; predictability; breeding vectors; data assimilation; numerical modelling; amplitude vacillation; structural vacillation; hysteresis

Received 22 May 2015; Revised 29 September 2015; Accepted 19 October 2015; Published online in Wiley Online Library

1. Introduction

Over a hundred years ago, early meteorologists realised that, while it might be possible to use the equations of atmospheric dynamics to determine future weather, they were beyond the capacity of mathematical analysis to solve. Instead, they constructed laboratory experiments that replicated the major influences on the Earth's atmosphere – gravity, heat, and rotation – arguing that recreating the atmosphere in the laboratory would allow it to be more readily understood and predicted (Vettin, 1857; Thomson, 1892; Bigelow, 1902; Abbe, 1907). Under certain conditions of experimental geometry and fluid, the principle of dynamical similarity could be used to ensure the same balance of forces in the atmosphere and the laboratory, and hence the two systems obeyed the same governing equations (Fultz, 1951).

Laboratory systems are still important for studying the basic mechanics underlying atmospheric motion. Such an environment is under full control of the experimenter, making the dynamics as repeatable as possible and hence subject to rigorous experiment. The thermally driven rotating annulus (Hide, 1953; Hide and Mason, 1975) is one such laboratory system (Figure 1(a)). It consists of two concentric cylinders maintained at different temperatures, rotating about a vertical axis, with fluid between them. The external forcing mimics the effects of rotation, gravity, and the difference in intensity of solar heating between low and

high latitudes on the midlatitudes of a planetary atmosphere (Figure 1(b)), with other effects rendered secondary. It is now firmly established as an insightful laboratory analogue for certain kinds of atmospheric dynamical behaviour.

The application of methods used to study and predict atmospheric circulation in one context may provide insights into their use in another. This article uses meteorological techniques to study the predictability of the rotating annulus experiment. The annulus displays a rich range of behaviour worthy of study in its own right. Like the atmosphere, the annulus experiment can display chaotic dynamics (Lorenz, 1963; Read *et al.*, 1992; Young and Read, 2008b). Furthermore, a large region of parameter space can be explored, so the experiment is useful for studying situations not just relevant to the Earth but to other planets and atmospheric predictability in general. For example, the Martian atmosphere is surprisingly predictable at certain times of year (Newman *et al.*, 2004; Rogberg *et al.*, 2010). The main aim of this article is to investigate the breakdown in predictability of baroclinic flow under a range of conditions in the laboratory – for how long can such flows in the annulus be forecast accurately, and what mechanisms cause predictability in annulus forecasts to be lost over time?

Lorenz (1975) defined two kinds of predictability: the first concerns the evolution of a system from initial conditions (the 'weather' problem), and the second the predicted behaviour given certain boundary conditions (the 'climate' problem). The second

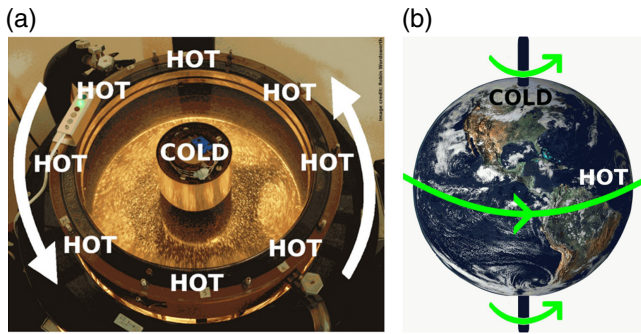


Figure 1. (a) Photograph of an annulus in AOPP's fluid dynamics laboratory (photo: Robin Wordsworth). (b) Illustration of the planetary analogue (image: NASA).

kind is well characterised in some regions of parameter space for the rotating annulus (Hide and Mason, 1975; Read *et al.*, 1992; Früh and Read, 1997; Read *et al.*, 2015), but the first kind has not previously been studied in much depth. Limited previous work includes using a radial basis function model to measure ι -shadowing times from experimental temperature data (Gilmour, 1998), and calculations of the Lyapunov exponents for various annulus flow regimes from experimental time series (Read *et al.*, 1992; Früh and Read, 1997), but these have only limited usefulness for characterizing the general predictability of complex systems.

The approach in this article is to adopt a similar framework to numerical weather prediction (NWP), combining a forecast model, a data assimilation scheme to calculate an accurate initial condition, and an ensemble generation mechanism to span the distribution of initial uncertainty. The Met Office/Oxford Rotating Annulus Laboratory Simulation (MORALS) will be used as the forecast model (Farnell and Plumb, 1976; Hignett *et al.*, 1985; Read *et al.*, 2000). For data assimilation we use analysis correction (Lorenc *et al.*, 1991), which we have already studied on its own in this context (Young and Read, 2013). We use the breeding method for ensemble generation (Toth and Kalnay, 1993, 1997). Breeding selects initial conditions based on the previous nonlinear evolution of the flow; small perturbations are periodically rescaled to 'breed' perturbations which identify the regions of most rapidly growing instability. We have also already applied breeding to the rotating annulus in isolation, using the perfect model scenario (PMS; Young and Read, 2008a).

By using these well-established techniques a second aim is to demonstrate that meteorological techniques can be applied in the controllable and reproducible environment of the rotating annulus in the laboratory and give sensible results about its predictability. There may be potential for annulus experiments to be used to study and evaluate new and proposed methods for prediction and assimilation that have not yet been put into operational use. The laboratory scale is a natural bridge between meteorologically relevant low-dimensional systems such as the Lorenz (1963) equations and the atmosphere itself.

In section 2 we present the full forecasting framework. Section 3 describes the forecasts, main results, and measures of predictability. In section 4 we explore in detail the wavenumber transitions seen in some forecasts, and in section 5 we conclude.

2. Methods

2.1. The rotating annulus model

MORALS solves the Navier–Stokes, continuity, and heat equations for temperature T and radial u , azimuthal v , and vertical w velocities in cylindrical polar coordinates (R, ϕ, z) , under the Boussinesq approximation. There is a diagnostic equation for kinetic pressure $\Pi = p/\rho_0$ solved using a Poisson equation (we shall refer to this quantity simply as 'pressure' from here). It uses an equation of state for density ρ and two constitutive relations for viscosity ν and thermal diffusivity κ , all quadratic functions of temperature. The full equations are listed by Young and Read (2013). The model domain consists of a fluid bounded by conducting cylinders at radii $R = a, b$ at temperatures T_a and T_b respectively ($\Delta T = T_b - T_a$), and insulating top and bottom boundaries at heights $z = 0, d$. The boundaries are no-slip, and the system rotates at angular velocity Ω about its central axis. The model grid is uniform in ϕ but stretched in R and z to resolve the boundary layers (three points in each). The resolution of the model is $24 \times 64 \times 24$ grid points on an Arakawa-C grid.

2.2. The forecasting framework

The Framework for Annulus Numerical Forecasting and Assimilation using Random Ensembles (FANFARE) has been developed to produce ensemble forecasts of the laboratory rotating annulus. Figure 2 shows a block diagram with the basic components of the framework. The forecast itself begins at $t = t_2$, and the stages before t_2 are used to set up the ensemble and an accurate initial condition.

During the development of FANFARE, the ensemble prediction scheme (Figure 2, dark grey) was first developed and used in the PMS with artificial data generated by MORALS (Young and Read, 2008a). Second, the assimilation stage (Figure 2, light grey) was developed and used on its own with data from the laboratory annulus (Young and Read, 2013). The complete framework combines and updates these two components, adding components to produce and verify forecasts against laboratory data (Figure 2, black).

2.2.1. Model spin-up

The first step in each forecast was to initialise MORALS, running from $t = 0$ to t_0 to generate the background state for the first assimilation. t_0 was chosen so any transient oscillations from the initial perturbation to the axisymmetric state had decayed.

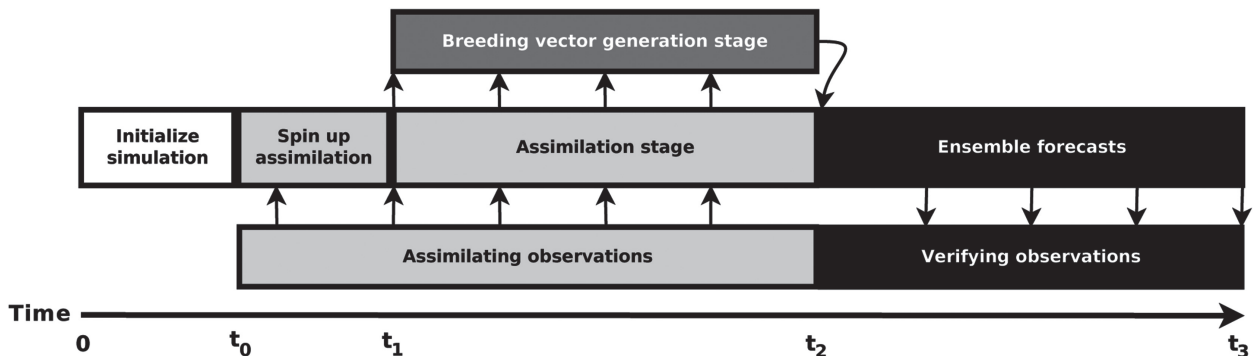


Figure 2. Block diagram showing the different stages and components of FANFARE. Arrows represent the flow of information between different stages. The stages are colour-coded: white, light grey, dark grey, and black for the four parts outlined in sections 2.2.1–2.2.4 respectively.

2.2.2. Data assimilation stage

The second stage of the forecast ran from t_0 to t_2 , the start of the main forecasts. This stage used data assimilation to produce a sequence of analyses \mathbf{x}^a which were then used as initial conditions for the breeding and forecast stages. Data assimilation combines observations, previous forecasts, climate, and relationships between physical quantities to produce the analysis. We used the analysis correction (AC) method (Lorenc *et al.*, 1991). Our implementation is fully described elsewhere (Young and Read, 2013), and the reader is referred to that article for full details. In AC the analysis is

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{W}\tilde{\mathbf{Q}} \left\{ \mathbf{y}^o - H(\mathbf{x}^b) \right\}, \quad (1)$$

where \mathbf{x}^b is a first guess (the *background*), \mathbf{y}^o is a set of observations, H maps the background to the observation points, and $\mathbf{W}\tilde{\mathbf{Q}}$ is a gain matrix. $\mathbf{W}\tilde{\mathbf{Q}}$ combines information about horizontal and vertical weighting functions between model and observation points, data density, time weighting, and observational, interpolation, and background-error covariances. AC is fairly straightforward to use and there was some local familiarity with it in the context of the Martian atmosphere (Montabone *et al.*, 2006). We did not use variational techniques because at the time tangent linear and adjoint models were not available for MORALS.

The assimilation stage had two purposes. The first was to provide initial conditions for a sequence of short forecasts used to generate breeding vectors between t_1 and t_2 . The second was to provide an initial condition for the main forecasts starting at t_2 . The parameters used in these assimilations were the same as in Table 3 of Young and Read (2013). Between t_0 and t_1 the assimilation ‘spun up’ before breeding started, calculating N_{before} analyses every Δt_{short} . This was required because the initial background state at t_0 and the laboratory observations are initially independent. The assimilation then ran from t_1 to t_2 , producing a series of analyses at each of the times required to start breeding cycles (the rescaling times, Δt_B apart). To avoid computational expense, the assimilations during the breeding stage were split into two sequences. All but the final N_{short} assimilations before each analysis required for breeding were separated by Δt_{long} . The final N_{short} assimilations were separated by Δt_{short} .

We chose N_{before} , Δt_{short} , and Δt_{long} so the analysis converged by the rescaling time, and also to optimise the available resources. The relevant time-scales here are the length of the assimilation window and the time between assimilations. Typically the second of these is as short as possible (Lorenc *et al.*, 1991, use one timestep), but that is not practical in our case. The assimilation window length for the atmosphere is typically a relevant mesoscale time-scale (say 6 h). In the annulus the comparable time-scale is the vacillation period, which is a couple of hundred seconds for regular flow, and around 50 s for chaotic flow; our assimilation windows are slightly shorter than this. Young and Read (2013) discuss these choices in some detail.

2.2.3. Breeding vector stage

The breeding method (Toth and Kalnay, 1993, 1997) is a powerful yet simple method for estimating the uncertainty in the analysis and for building an ensemble of initial conditions reflecting the nonlinear evolution of the flow. The result is a set of ‘breeding vectors’ (or BVs), which identify regions of most rapidly growing instability. The basic procedure is

- (i) perturb the analysis with random noise;
- (ii) forecast a short time into the future from the unperturbed and perturbed states;
- (iii) calculate the difference between the two;
- (iv) rescale the difference to the size of the original perturbation (if it has grown);

(v) use the rescaled difference (the BV) as the new perturbation; and

(vi) repeat from step (ii).

We chose to use breeding vectors here primarily for practical reasons. An alternative is to use the method of singular vectors (Buizza and Palmer, 1995), which selects initial conditions based on the direction of the fastest-growing linear perturbations to the analysis state over a short optimization period. However, BVs are computationally cheaper to construct than singular vectors, and at the time there was no tangent linear model for MORALS (although there is now,* so we may consider singular vectors in the future).

The BV generation stage begins at t_1 and ends at t_2 , and is used to generate an ensemble of forecast initial conditions by perturbing the analysis at t_2 . Multiple cycles are required for a coherent structure to develop in the BVs.

In this work we use ‘self-breeding of twin forecasts’ (Toth and Kalnay, 1997). This differs slightly from our PMS work (Young and Read, 2008a), which used standard breeding. Self-breeding runs M breeding cycles with identical positive and negative perturbations and bases the BV on the difference between these forecasts, rather than between perturbed and non-perturbed forecasts. This method was used operationally by the National Centers for Environmental Prediction (NCEP) during the 1990s (Toth and Kalnay, 1993).

The procedure starts at $t = t_1$ with analysis \mathbf{x}^a computed as described above. M initial Gaussian noise perturbations are produced in the \mathbf{u} and \mathbf{v} fields (MORALS solves a Poisson equation for pressure to ensure the flow remains non-divergent). Because only \mathbf{u} and \mathbf{v} observations are available the \mathbf{w} , \mathbf{T} , and $\mathbf{\Pi}$ fields are not perturbed at t_1 . This noise has mean zero and standard deviation a constant fraction (the ‘bred vector amplitude’, denoted F) of the climatological RMS variance of \mathbf{u} and \mathbf{v} at that point (denoted $\sigma_{\mathbf{u}^a}$ and $\sigma_{\mathbf{v}^a}$ in the two fields). F is determined empirically; see Table S2 and Figure S3 of Document S1 for more detail, along with the climate calculation (Table S1 and Figure S2, Document S1).

Our choice of F puts our experiments within the region of slow growth identified by Toth and Kalnay (1993). In fact, the MORALS error growth in the regimes studied seems to be independent of perturbation size; a similar analysis to Harlim *et al.* (2005, Figure 1) showed there is no faster growth at small perturbation size. Both these articles attributed this faster growth to fast convective modes, but we suspect these may not exist in our model, which is non-hydrostatic, has no subgrid-scale parametrizations, and shows no evidence of strong inertia-gravity waves (at the points in parameter space studied here, at least).

These M perturbations are added and subtracted from the analysis at t_1 to generate $2M$ perturbed initial conditions (superscript p):

$$\mathbf{u}^{p,m\pm}(t_1) = \mathbf{u}^a(t_1) \pm F \sigma_{\mathbf{u}^a} N(0, 1), \quad m \in [1, \dots, M], \quad (2)$$

and similarly for \mathbf{v} (the other fields take their analysis values). Each initial condition is then used to start MORALS, which integrates it forward by Δt_B (the ‘rescaling time’). At $t = t_1 + \Delta t_B$, the difference (superscript d) is calculated between pairs of forecasts started from positive and negative initial conditions (now all five fields):

$$\mathbf{x}^{d,m}(t_1 + \Delta t_B) = \mathbf{x}^{p,m+}(t_1 + \Delta t_B) - \mathbf{x}^{p,m-}(t_1 + \Delta t_B), \quad m \in [1, \dots, M], \quad (3)$$

to give a set of M difference fields. These differences are then rescaled. The norm, denoted $\eta^m(t)$ for perturbation m , is a

*U. Achatz, 2010; personal communication.

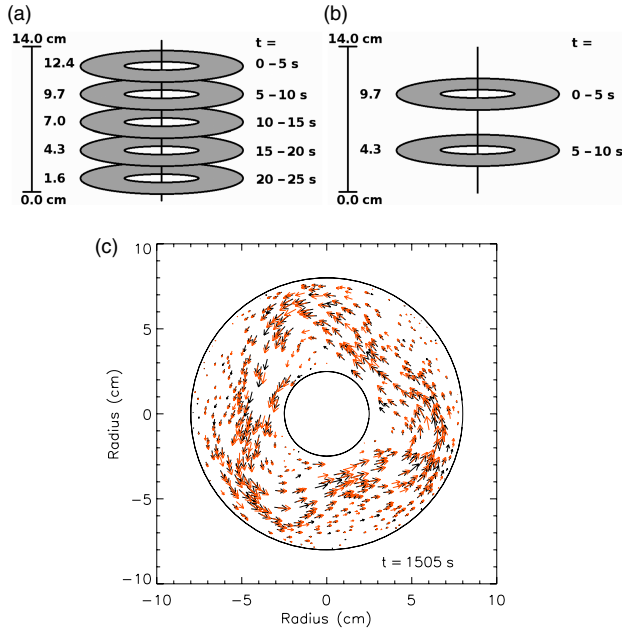


Figure 3. Laboratory rotating annulus observations. (a) Schematic showing the vertical distribution of annulus observations in datafiles expf1 and expf2 (2S/3AV). (b) As (a) but for expf5 and expf6 (3SV). (c) Example horizontal velocity observations in the expf2 datafile, showing a single dataset at $z = 9.7$ cm. The two independent subsets are shown as red (grey in print) and black vectors.

volume-weighted horizontal velocity magnitude:

$$\eta^m(t) = \sqrt{\int \int \left[\{ \mathbf{u}^{d,m}(t) \}^2 + \{ \mathbf{v}^{d,m}(t) \}^2 \right] dV} / \int dV. \quad (4)$$

The growth after cycle c is

$$g^m(t_1 + c\Delta t_B) = \frac{\frac{1}{2}\eta^m(t_1 + c\Delta t_B)}{\eta^m(t_1)}, \quad (5)$$

where $\eta^m(t_1)$ is calculated at the start of the first breeding cycle using the difference between perturbation and analysis. The $1/2$ in the numerator is there because subsequent η are calculated from the difference between forecasts from positive and negative perturbations. A growing perturbation has $g^m(t_1 + c\Delta t_B) > 1$, and if the perturbation has grown then all five difference fields are scaled by this ratio to produce BVs:

$$\mathbf{x}^{bv,m}(t_1 + c\Delta t_B) = \frac{\mathbf{x}^{d,m}(t_1 + c\Delta t_B)}{g^m(t_1 + c\Delta t_B)}. \quad (6)$$

A decaying perturbation has $g^m(t_1 + c\Delta t_B) \leq 1$, in which case the fields are not rescaled and the BV is simply half the difference field:

$$\mathbf{x}^{bv,m}(t_1 + c\Delta t_B) = \frac{1}{2}\mathbf{x}^{d,m}(t_1 + c\Delta t_B). \quad (7)$$

The M new bred vectors then perturb the next analysis at $t = t_1 + c\Delta t_B$ (note all five fields are now perturbed):

$$\mathbf{x}^{p,m\pm}(t_1 + c\Delta t_B) = \mathbf{x}^a(t_1 + c\Delta t_B) \pm \mathbf{x}^{bv,m}(t_1 + c\Delta t_B). \quad (8)$$

These $2M$ initial conditions are then integrated forward by MORALS. This cycle is repeated in intervals of Δt_B , until t_2 .

Corazza *et al.* (2003) showed that BVs can be used to estimate the background-error covariance (the ‘errors of the day’). FANFARE was designed so no information passed from the breeding to the assimilation, as it made the scheme easier to code in modular form because the complete sequence of

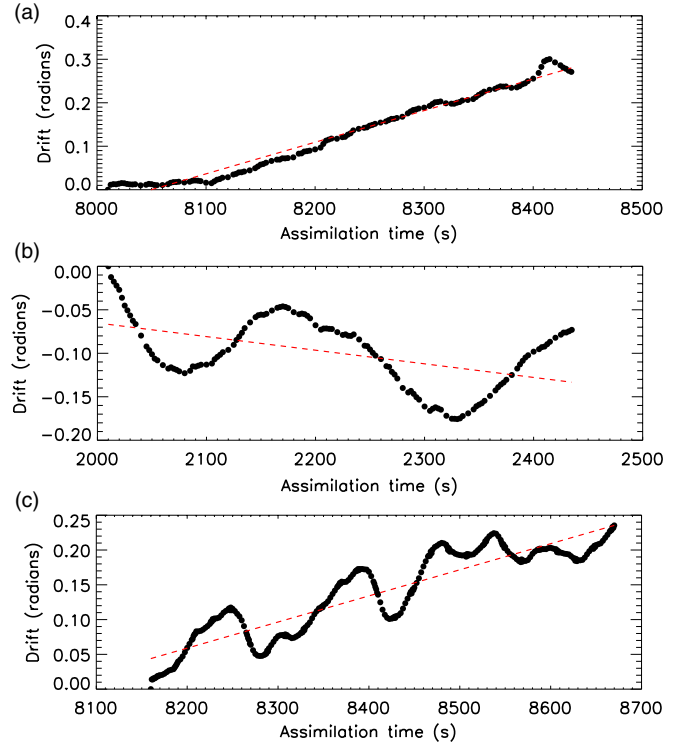


Figure 4. Wave drift during the assimilation stage for representative forecasts in the three flow regimes. Dots are wave drift for each analysis, and the dashed line is the linear best fit. (a) forecast A2 (2S), (b) forecast C2 (3AV), and (c) forecast G1 (3SV).

analyses could be processed before any BVs. However, no updated information about the background-error covariance can pass from the breeding to the assimilation, so we used the same background-error covariance for each analysis. The initial background to experimental error ratio was sufficiently large for this approximation to be reasonable in almost all cases.

2.2.4. Forecast stage

In the forecast stage the ensemble of initial conditions at t_2 was used to predict subsequent laboratory observations up to t_3 . The ensemble contained an unperturbed control forecast from the assimilated analysis, and a set of $2M$ perturbed forecasts created using the BVs:

$$\left. \begin{aligned} \mathbf{x}^c(t_2) &= \mathbf{x}^a(t_2), \\ \mathbf{x}^{p,m\pm}(t_2) &= \mathbf{x}^a(t_2) \pm \mathbf{x}^{bv,m}(t_2), \quad m \in [1, \dots, M]. \end{aligned} \right\} \quad (9)$$

These $2M + 1$ initial conditions were integrated by MORALS from t_2 to t_3 , and then compared with laboratory observations at 5 s intervals between those times.

2.3. Laboratory data

Four laboratory *datafiles* were used in this work, from the AOPP archive of rotating annulus data. Each contains about 10 000 s of horizontal velocity observations (\mathbf{u} and \mathbf{v}) grouped in *datasets* every 5 s. Three datafiles (expf2, expf5, and expf6) were used for forecasting, and the fourth (expf1) was just used to compute climate distributions. expf1 and expf2 contain flow in regular flow regimes 2S (steady) and 3AV (amplitude vacillating) (Hide and Mason, 1975), and expf5 and expf6 are in the 3SV (structurally vacillating) chaotic flow regime (Read *et al.*, 1992).

The apparatus used allows measurements to be taken at up to five different levels in the annulus, shown in Figure 3(a,b). Flow velocities are calculated using Particle Imaging Velocimetry over a 1 s interval. The random error in each velocity component was

estimated (Young and Read, 2013) to be 0.0057 cm s^{-1} , which is much smaller than the velocity variability. The sequence of observations is staggered: the first 5 s take data at the top level, followed by 5 s at each of the other heights in turn, then the cycle repeats. There are no observations of \mathbf{w} , \mathbf{T} , or Π . Figure 3(c) shows an example velocity field dataset from the expf2 datafile.

Each dataset is split into two or more *subsets*; in Figure 3(c) these are shown in red and black. All datafiles have two subsets except expf5, which has three. The subsets are separated by about a second, and while some measurements may be the same particle measured more than once, the subsets are sufficiently uncorrelated to be used as independent data valid at the same time (the separation time is much smaller than the shortest time-scale of the flow). The forecasts use one subset for assimilation and breeding and to generate the initial condition, and verify against another, independent, subset.

Full details of the expf2, expf5, and expf6 datafiles, including how the data were preprocessed and filtered, and how the random error in each velocity vector was calculated, are provided in Young and Read (2013), and the reader is referred to that article for details. expf1 was not used in that article and so equivalent details are included in the Document S1 (Figure S1). The data can be downloaded from ORA-Data (Young *et al.*, 2015).

2.4. Wave drift

A correction is made to the completed forecasts before they are verified against observations. In initial testing wave drift was identified as a major source of forecast error. As the forecast and the real annulus evolve over time, the waves in the two can precess around the tank at different rates, because a wave's phase speed is sensitive to a number of parameters which must be matched between forecast and experiment, particularly ΔT and Ω . Systematic errors in the model make this harder than just setting the same parameters as were used in the laboratory. Young and Read (2008b) identified a systematic shift in Ω of -0.11 rad s^{-1} for regime transitions at low Ω between MORALS and laboratory observations. This shift is accounted for, but the estimate is only accurate to $\pm 0.02 \text{ rad s}^{-1}$ and only at low Ω where transitions are sharp.

In practice the wave drift can be corrected empirically. We estimate the drift before the forecast begins by comparing an unassimilated run with the analysis during the data assimilation stage. The drift is the difference between the phases of the dominant pressure modes at mid- (R, z) in the two cases. The average drift rate is calculated over the assimilation stage (Figure 4), and then a reverse azimuthal transformation is applied to the model fields. We use this to correct both the mean forecasts, assuming constant drift into the future, and also to correct the BVs, as both the positive and negative perturbations are subject to wave drift relative to observations, but the subsequent BV is added to an analysis not subject to the drift.

3. Forecasts of the laboratory annulus

Fourteen forecasts were run, each covering a sequence of observations at a single rotation rate Ω . To be suitable, the sequence of observations at a particular Ω had to meet the following conditions:

- (i) Ignoring the first 200 s at each Ω to allow the flow to equilibrate, there must be enough time before the next Ω transition to generate BVs (400–500 s) and to run a forecast of at least 1000 s. This excluded all of datafile expf1, so this was only used to calculate climate statistics.
- (ii) Observation times with no data should be avoided during the assimilation and BV generation stages.
- (iii) Ω values where there are wavenumber transitions in the data are avoided (e.g. $\Omega = 0.8 \text{ rad s}^{-1}$ in expf2).

- (iv) Sequences used to test the breeding or assimilation parameters should ideally be avoided, although this was not possible in practice with the 2S forecasts.

Seven Ω values satisfied these conditions. One is in the 2S regime, two are in 3AV (the 'regular regimes'), and four are in 3SV (the 'chaotic regime'). Figure 5 shows the position of each forecast in its datafile, and Table 1 lists the forecast parameters. The forecasts all fell on the line $(Ta, \Theta_T) = (4.89 \times 10^6 \Omega^2, 0.125 \Delta T / \Omega^2)$ in the standard annulus regime diagram, where Ta and Θ_T are the Taylor and thermal Rossby numbers respectively. The tank had inner and outer radii $a = 2.5 \text{ cm}$ and $b = 8.0 \text{ cm}$ respectively, depth $d = 14.0 \text{ cm}$, with a 83% water/17% glycerol working fluid by volume (Prandtl number 13.4). ΔT is approximately 4.05 K in datafile expf2, and 4.02 K in expf5 and expf6.

Other parameters common to each forecast (Table S2, Document S1) were the long time-scale between analyses in the breeding stage $\Delta t_{\text{long}} = 5.0 \text{ s}$, the number of assimilation cycles before the breeding stage $N_{\text{before}} = 10$, the bred vector amplitude $F = 0.2$, and the breeding rescaling time $\Delta t_B = 20 \text{ s}$. Forecasts in regimes 2S and 3AV had the time between analyses in the assimilation stage $\Delta t_{\text{short}} = 2.5 \text{ s}$ and $N_B = 20$ breeding vector cycles, while these were $\Delta t_{\text{short}} = 1.0 \text{ s}$ and $N_B = 25$ in regime 3SV. The model time step was $\delta t = 0.02 \text{ s}$ for the regular regimes and $\delta t = 0.01 \text{ s}$ for the chaotic regime.

We spent some time investigating the optimal breeding parameters; see Document S1 and Young and Read (2008a) for details. The rescaling time needs to be sufficiently short for baroclinic instabilities not to saturate (Toth and Kalnay, 1993, their Figure 6). Hide and Mason (1975, their Figure 15) shows the time-scale for growth of the most unstable baroclinic mode for a typical rotating annulus; our experiments fall in the 30–100 s range, while for the Earth's atmosphere it is about 4 days (Vallis, 2006, his Eq. 6.98). Both in our case and for atmospheric breeding a handful of rescaling times cover the period it takes a baroclinic wave to grow.

Each forecast computed eight independent breeding cycles for a forecast ensemble of 16 perturbed forecasts plus the control forecast. Two forecasts were done at each Ω , for repeatability of the experiment in each case, and they are identical except a few details (Table 1).

It should be noted that to produce analyses between t_0 and t_2 , AC uses observations from $t_0 - t^f$ to $t_2 + t^b$, where t^f and t^b are the lengths of the assimilation window forwards and backwards in time from a single observation (Young and Read, 2013), so the initial conditions for the main forecasts use some information from the future. We use t_2 as the zero point for the forecast time axes, so it should be noted in advance that the 'real' forecast only begins at $t_2 + t^b$. In practice t^b is small compared with most predictability times (26 and 21 s for the regular and chaotic regimes respectively).

3.1. Breeding vectors

The breeding vectors produced in the different flow regimes highlight interesting differences between the flows and reflect the unstable nature of some of the flow regimes.

In Figure 6 we show all the azimuthal velocity BVs at the start of the 3AV C2 and 3SV D2 forecasts. In the regular regimes the BVs generally resemble waves of the same wavenumber as the analysis, although not necessarily in phase. In some cases wavenumber-2 BVs form in the 3AV forecasts. By the end of the breeding stage, a coherent structure has formed in each perturbation, with the BVs for a particular forecast generally all visually similar to each other. For a given feature, some BVs have a peak and others a trough, which is expected from the symmetry of the initial perturbation. In the chaotic regime, the BVs do not form clear wave structures. Over the course of the breeding stage, features grow and decay as the

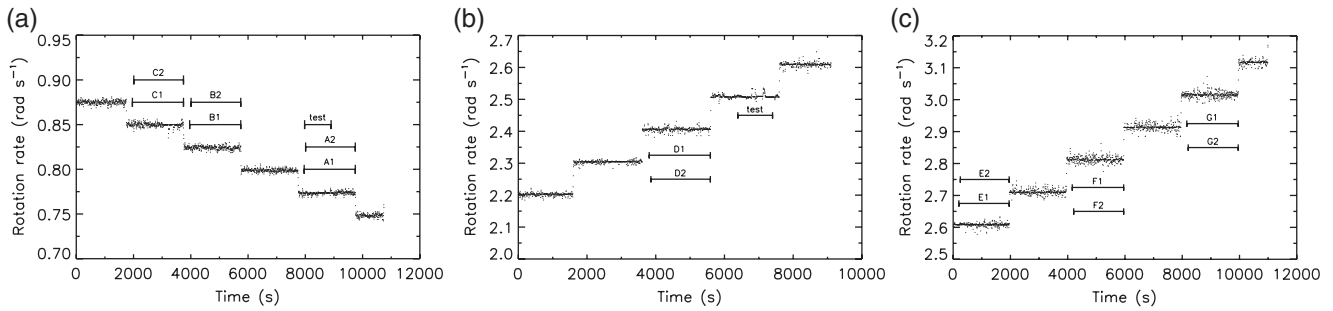


Figure 5. Positions of the forecasts within the laboratory datafiles. (a) expf2: forecasts A, B, and C, (b) expf5: D, and (c) expf6: E, F, and G.

Table 1. Parameters used for the main forecasts. Forecasts with suffix ‘1’ have a 2000 s background run, assimilation dataset 1 and verification dataset 2. Those with suffix ‘2’ have a 2050 s background, using assimilation dataset 2 and verification dataset 1.

Flow regime	Run	Datafile	Rotation rate Ω (rad s ⁻¹)	Assimilation start t_0 (s)	Breeding start t_1 (s)	Breeding length $t_2 - t_1$ (s)	Forecast start t_2 (s)	Forecast length $t_3 - t_2$ (s)	Forecast end t_3 (s)
2S	A1	expf2	0.775	7960	7985	400	8385	1360	9745
2S	A2	expf2	0.775	8010	8035	400	8435	1310	9745
3AV	B1	expf2	0.825	3960	3985	400	4385	1360	5745
3AV	B2	expf2	0.825	4010	4035	400	4435	1310	5745
3AV	C1	expf2	0.850	1960	1985	400	2385	1360	3745
3AV	C2	expf2	0.850	2010	2035	400	2435	1310	3745
3SV	D1	expf5	2.4	3810	3820	500	4320	1270	5590
3SV	D2	expf5	2.4	3860	3870	500	4370	1220	5590
3SV	E1	expf6	2.6	200	210	500	710	1245	1955
3SV	E2	expf6	2.6	250	260	500	760	1195	1955
3SV	F1	expf6	2.8	4160	4170	500	4670	1285	5955
3SV	F2	expf6	2.8	4210	4220	500	4720	1235	5955
3SV	G1	expf6	3.0	8160	8170	500	8670	1285	9955
3SV	G2	expf6	3.0	8210	8220	500	8720	1235	9955

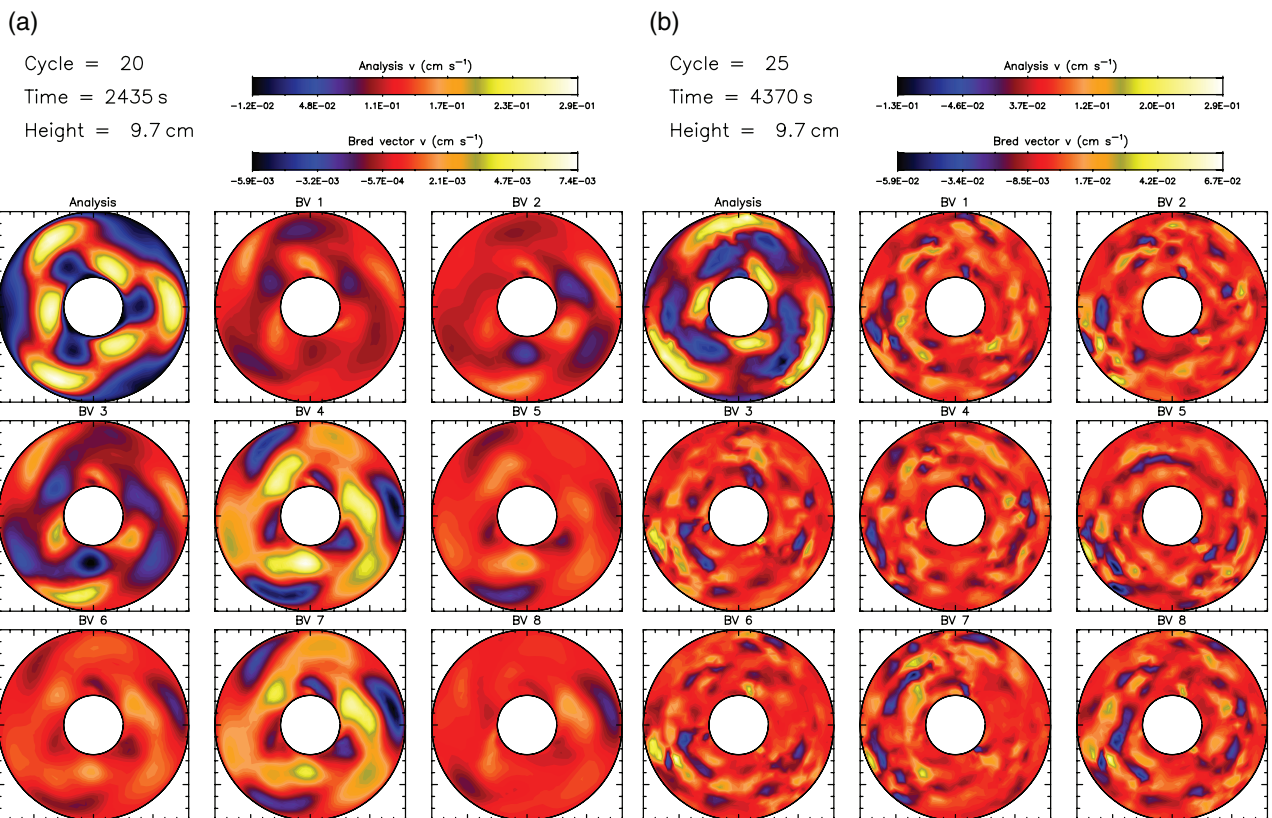


Figure 6. Azimuthal velocity breeding vectors for forecasts (a) C2 (3AV) and (b) D2 (3SV), at the start of the forecast ($t = t_2$). The analysis at $z = 9.7$ cm is shown in the top left, followed by the eight BVs generated by rescaling the difference between positive and negative forecasts. Figure S12 of Document S1 shows equivalent plots for the other forecasts.

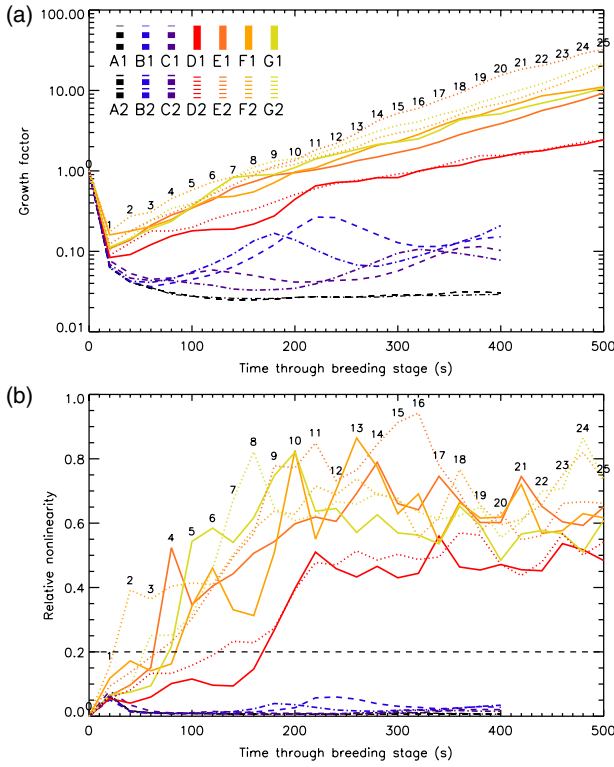


Figure 7. Diagnostics from the BV generation stage. Numbers show the breeding cycle at that time. (a) Cumulative growth factor (Eq. (10)). Each line shows the (geometric) mean of the growth factors over all BVs in the ensemble. (b) Relative nonlinearity (Eq. (11)). Each line shows the (arithmetic) mean of the R_{NL} values for each bred-pair in the ensemble. The dashed line shows $R_{NL} = 0.2$.

underlying structure of the analysis changes, on time-scales up to 100 s.

Figure 7 shows the cumulative growth for BVs in all forecasts during the BV generation stage. The cumulative growth of a BV after c breeding cycles is

$$G_c = g_c \prod_{\substack{i=1 \\ g_i > 1}}^{c-1} g_i, \quad (10)$$

using g from Eq. (5). Essentially G_c is the growth in the current step multiplied by the cumulative growth up to the last time the fields were rescaled. This formulation is required because the BV is only rescaled if it grows during the cycle. The regular regime BVs decay (black, blues) and the chaotic regime BVs grow (reds, greens). All perturbations decay in the first cycle because all directions in the phase space are perturbed by the Gaussian perturbation but most are stable, decaying directions. Only in later cycles do unstable directions become dominant. In the regular regimes, after the initial decay the growth is slow over cycles 3 \rightarrow 20, at a multiplicative rate of $\sim 1.07 \text{ cycle}^{-1}$. This growth is non-exponential, with a BV doubling time around 200 s. 3SV growth is faster, at a factor 1.23 cycle^{-1} over cycles 2 \rightarrow 25, and is exponential, with a BV doubling time around 70 s.

Another relevant diagnostic is the BV relative nonlinearity (Gilmour *et al.*, 2001), defined as

$$R_{NL}(t) = \frac{||\mathbf{x}^{p,m+}(t) + \mathbf{x}^{p,m-}(t)||}{\frac{1}{2} (||\mathbf{x}^{p,m+}(t)|| + ||\mathbf{x}^{p,m-}(t)||)}, \quad (11)$$

using the norm in Eq. (4). This measures the nonlinearity of the flow by comparing the evolution of positive and negative perturbations. A BV is nonlinear if the sufficient condition $R_{NL} \geq 0.2$ is met (Gilmour *et al.*, 2001). Figure 7 shows R_{NL} for each forecast. The difference between regular and chaotic regimes here is clear. The regular regime forecasts are consistently below $R_{NL} = 0.2$, and the chaotic regime forecasts all exceed this

value within the first ten cycles. The only general trend among the SV forecasts is that the relative nonlinearity of F and G (higher Ω) is generally higher than for D and E.

Finally, some authors (Toth and Kalnay, 1993; Houtekamer and Derome, 1994; Corazza *et al.*, 2003) argue that BVs are representative of analysis uncertainty because of the similarities between the analysis cycle and the breeding cycle. We looked for positive correlations between the spread of analysis values generated using our data assimilation scheme and the spread of BVs at the same time. The correlations were not particularly strong, but the spatial distributions were sufficiently similar to support a connection. The BV spread was typically an order of magnitude smaller than the analysis error for 2S and 3AV, and was comparable for 3SV.

3.2. Main forecasts: general observations

3.2.1. Steady waves: forecast A

The two forecasts in the steady regime 2S are generally good, with the large-scale wave structure maintained throughout the forecast. Forecasts A1 and A2 represent the best predictability that can be hoped for in this work.

At a single point the observations are well-predicted, shown in Figure 8(a). There are oscillations at $t \sim 8400$ s and $t \sim 9400$ s, which appear to be inertia-gravity waves close to the inner cylinder. They do not obviously appear in the observations, but this is not particularly surprising as their presence in model runs is sensitive to the forcing parameters. These features may be generated by a boundary-layer instability (Jacoby *et al.*, 2011).

Forecast A2 is shown using spaghetti plots in Figure 9. These show the ensemble following the observations closely until the

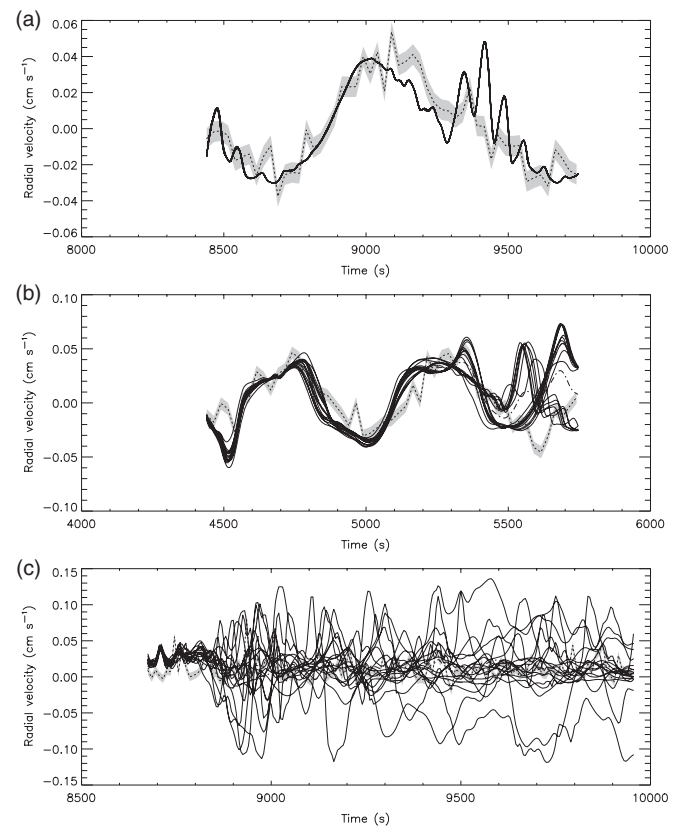


Figure 8. Time series showing the radial velocity at $R = 3.0 \text{ cm}$, $\phi = 0.0 \text{ rad}$, $z = 4.3 \text{ cm}$. This point is in the jet stream near the inner part of a cyclone in the 3SV observations. The ensemble forecast values are shown by solid lines, and the observed values (interpolated to the same point) are shown by dotted lines with the approximate error shaded. (a) Forecast A2 (2S), (b) forecast B2 (3AV), and (c) forecast G1 (3SV).

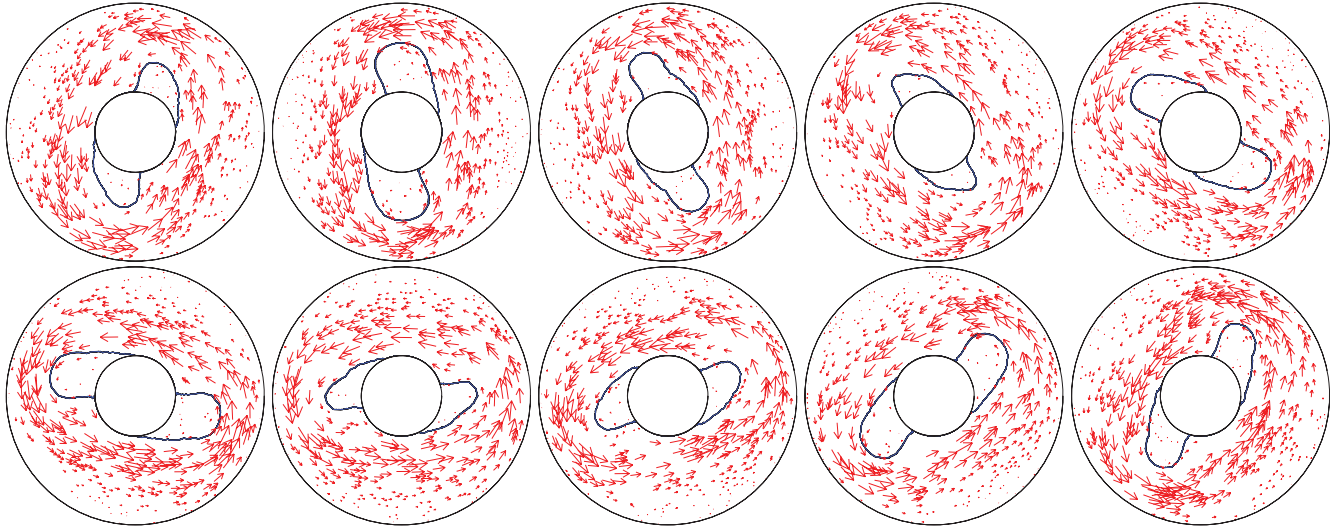


Figure 9. Sequence of spaghetti plots showing forecast A2 (2S, $\Omega = 0.775 \text{ rad s}^{-1}$) at $z = 9.7 \text{ cm}$. Frames are separated by 100 s, and the final frame is at $t_2 + 970 \text{ s}$. Each frame shows the $\Psi = 0$ streamfunction contour for each of the positive ensemble members (black solid), negative ensemble members (black dotted), control forecast (brown solid), ensemble mean (blue solid), and observed vector velocity field (red arrows). Note that in this instance all these lines are on top of each other – this is not the case in later examples.

end of the forecast, with an accurate drift rate relative to the observations. The forecast spread about the ensemble mean is small; there is probably no advantage in using an ensemble in this regime. The ensemble mean \bar{x}^B is the average of the forecasts in the ensemble (excluding the control forecast), and the forecast spread $\sigma_{\bar{x}}^B$ is the standard deviation of values around it.

Plots of the normalised root mean square error (RMSE) are shown in Figure 10(a). The RMSE measures forecast accuracy, using the difference between forecast and observations, and this difference is normalised by the climatological RMS variance, at each observation location. For a forecast of N forecast–verification pairs $[f_n, o_n]$ each with weight w_n and observational climatological RMS variance c_n , the RMSE is $\sqrt{\sum_{n=1}^N w_n \{(f_n - o_n)/c_n\}^2}$. We use horizontal velocity magnitudes (i.e. ‘wind speed’) for this calculation, as this uses all the observational information. Our observations are distributed non-uniformly in space, so we assign a normalised weight w_n to each pair based on a simple estimate of local observation density (Young and Read, 2013). The RMSE for this regime shows that it may be predictable for much longer than the available data, as the RMSE is not increasing. If these forecasts were longer, predictability would eventually be lost by the forecast and observations going out of phase.

Tracking individual features in the flow can be used to decompose the RMSE into components representing displacement, amplitude, and residual contributions (Hoffman *et al.*, 1995). In the annulus, these correspond to error due to (i) the azimuthal displacement of the large scale wave relative to observations;

- (ii) error in the amplitude of the large-scale wave; and
- (iii) a residual term due to small-scale features.

These diagnostics are particularly interesting as they reveal the error due to a poor drift rate calculation, and the error left once the position and shape of the wave have been corrected; this represents the ‘true’ error in the model, which cannot be removed by simple transformations.

We calculated this decomposition using the control forecasts. The forecast was rotated by a range of azimuthal angles to find the rotation minimising the horizontal velocity magnitude RMSE. Second, the amplitude of the rotated field was multiplied by values in the range 0–4 to find the minimum RMSE. This gave three model fields (forecast, displaced, and amplified), which were used to calculate the displacement (forecast–displaced), amplitude (displaced–amplified), and residual errors

(amplified–observations) by calculating the RMSE between the different fields.

Figure 11(a) shows the RMSE decomposition, again normalised by the climatological RMS variance. The accurate drift rate for the 2S forecasts is confirmed; only a small fraction of the total RMSE is attributed to displacement error, while most is attributed to residual error. There is a small fraction attributed to amplitude error, because in the forecast there is a small vacillation in the dominant wave which does not appear in the observations.

A different approach to forecast verification is to evaluate forecast skill, which quantifies the accuracy of a probability forecast against a reference (Woodcock, 1976). We use the Brier skill score (Brier, 1950; Wilks, 2001) here:

$$\text{BSS} = \frac{1 - \text{BS}_f}{\text{BS}_c + \frac{1}{M} p_1 (1 - p_1)},$$

where BS_f and BS_c are the Brier scores for ensemble and climatological reference forecasts respectively. The Brier score is a RMS probability forecast error for yes/no outcomes (e.g. rain/no rain, or wind speed above a certain value), and is defined for N forecast–observation pairs $[p_n, d_n]$ with weights w_n as $\text{BS} = \sum_{n=1}^N w_n (p_n - d_n)^2$. p_n is calculated from the ensemble or, in the climatological reference case, from the climate distribution of a particular quantity, and d_n is either zero or one depending on whether the event occurs or not. Our BSS includes an additional bias correction term (Weigel *et al.*, 2007). (M is the size of the ensemble, and p_1 is the climate probability of the ‘yes’ outcome.) In the annulus context any ‘event’ is arbitrary, but a threshold on the horizontal velocity magnitude seems natural, so we predict whether the observations exceed the 60th percentile of the horizontal velocity magnitude climate distribution at each point in the tank.

The predictability limit for a forecast based on skill is defined as the point when the BSS falls to zero. We estimated 95% confidence intervals for our BSS values using bootstrapping (Mason, 2008). The forecast–observation pairs used to calculate the BSS were randomly resampled, with replacement, to get a sample the same size as the original dataset. The sample score was then calculated and this was repeated 1000 times, giving a distribution of values from which the 95% interval was extracted. We defined the predictability limit as the first time when the 95% confidence interval included zero. Figure 12(a) shows the BSS. It confirms that in the 2S forecasts skill relative to the climate forecast was retained until the end of the forecast, consistent with the RMSE.

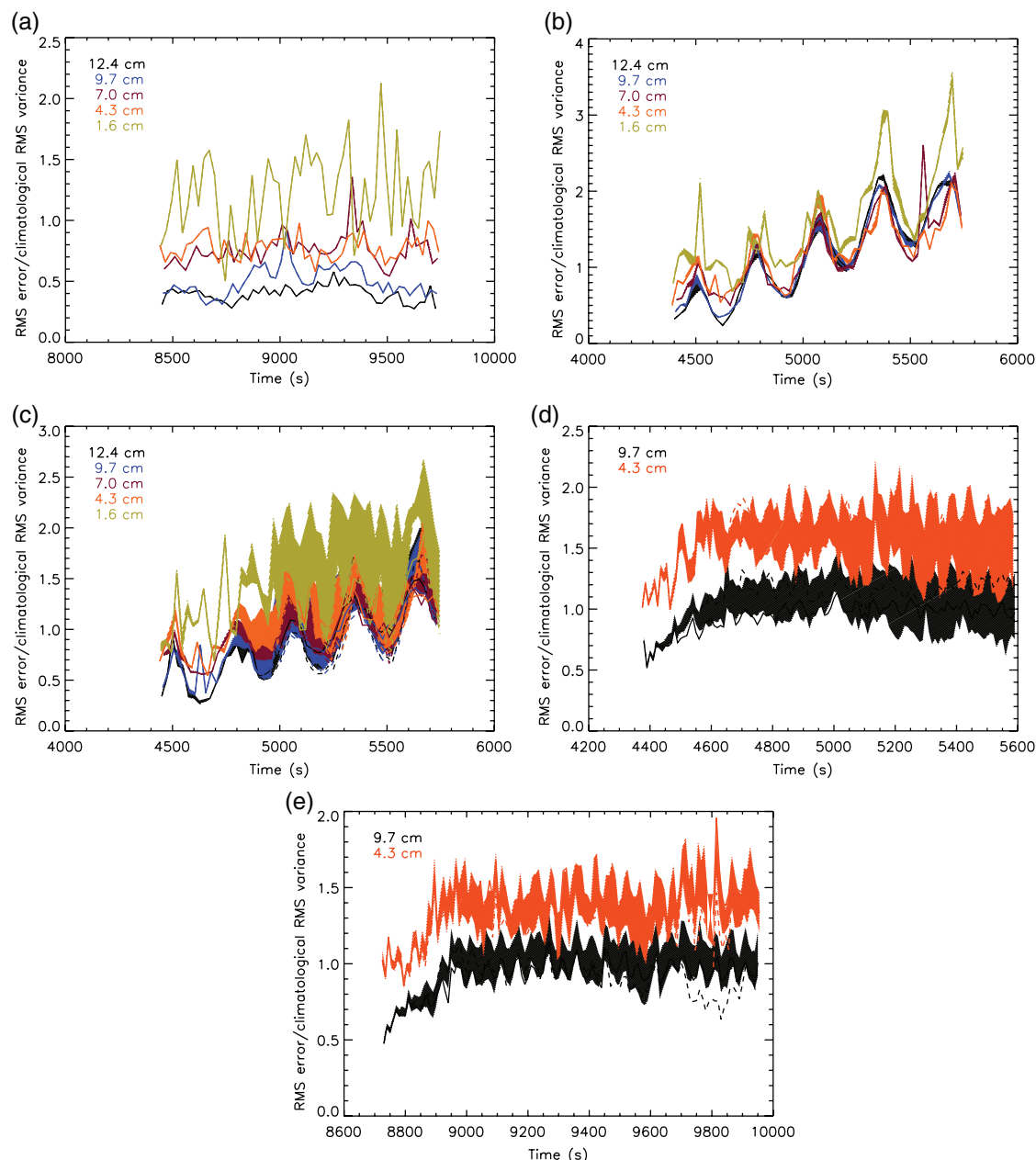


Figure 10. Time series of RMSE using the horizontal velocity magnitude field for representative forecasts in the three regimes. The different vertical levels are shown in different colours. The shaded regions show the 10–90th percentiles of each ensemble, solid lines are ensemble means, and dashed lines are control forecasts. The errors are normalised by the climatological RMS variance at each observation point. Absolute RMS errors are typically $0.02\text{--}0.2\text{ cm s}^{-1}$. (a) Forecast A2 (2S), (b) B1 (3AV), (c) B2 (3AV), (d) D2 (3SV), and (e) G2 (3SV).

3.2.2. Amplitude vacillation: forecasts B and C

These four forecasts are in the 3AV regime close to the transition between the 3AV and the 2S regimes, and have the wave's vacillation to predict in addition to its position. Spaghetti plots for forecast B2 are shown in Figure 13, and individual ensemble member pressure fields are shown for a point during forecast B2 in Figure 14(a). Predictability is lost in the 3AV forecasts by two main mechanisms.

First, the estimate of wave drift in the forecast compared with the observations is generally poor, which leads to a loss of forecast quality as a result of *forecast drift*. B1 and C1 lose predictability by this mechanism, and the reason this drift is poorly estimated was discussed in the previous section.

Second, in forecasts B2 and C2 predictability is lost primarily by an azimuthal *wavenumber transition* from $m = 3 \rightarrow 2$. This was not expected. An example is shown in Figure 15, where the dominant wave changes during the forecast to $m = 2$. In B2 seven of the 16 perturbed ensemble members transition to $m = 2$, and in C2 all 16 perturbed ensemble members (plus

the control forecast) transition to the lower wavenumber. A number of reasons why these transitions occur are discussed in section 4.

Figure 8(b) shows the evolution of the B2 ensemble at a single point. The transition is around $t = 5200\text{ s}$ as the ensemble splits into two parts for the $m = 3$ and $m = 2$ ensemble members. In the equivalent plot for C2 (not shown), the transition is not as obvious as in B2, because in that case all ensemble members transitioned. While B1 and C1 go out of phase over time in such a way that could be corrected by a simple azimuthal translation, this is not the case in C2, where the wave's fundamental nature has changed.

RMSE diagnostics are shown in Figure 10(b,c), showing one case (B1) where predictability is lost via forecast drift, and another (B2) where it is lost primarily by a wavenumber transition. In B1, where there is no transition, the spread of RMSE is smaller than in B2. Overall, the RMSE in the 3AV forecasts is higher than in 2S, and it continues to increase over the course of the forecasts. There is an oscillation on the vacillation time-scale caused by the mismatch in vacillation periods between forecasts

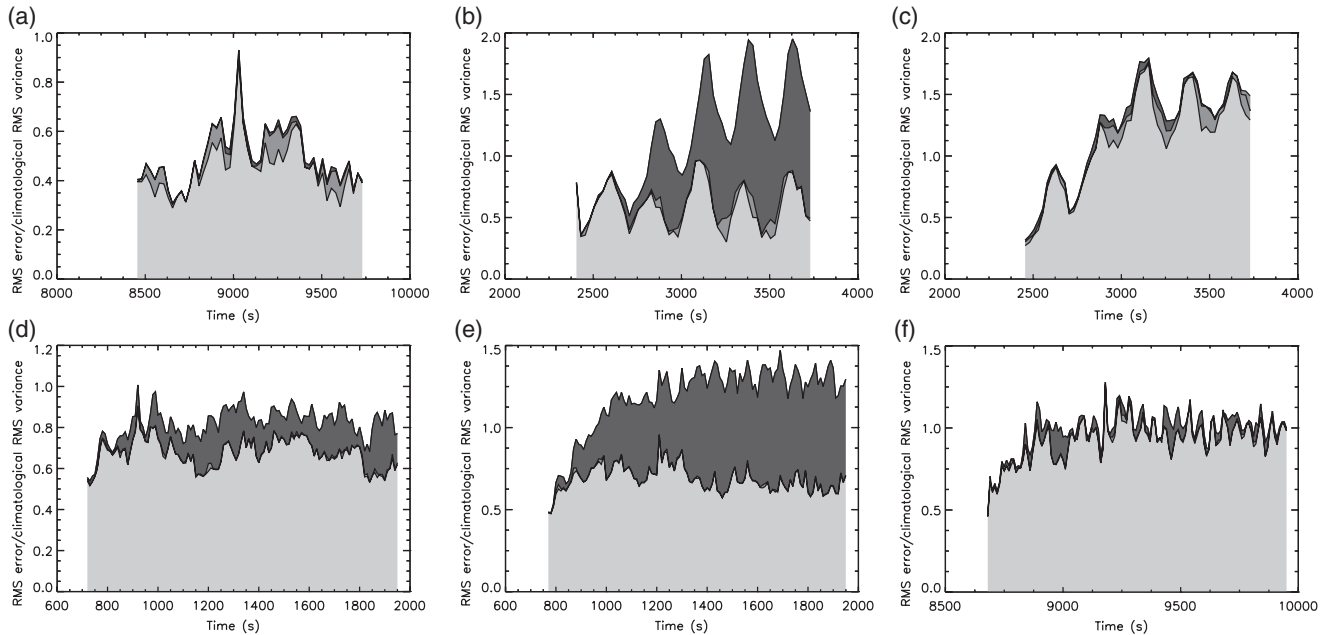


Figure 11. Decomposition of the RMSE at $z = 9.7$ cm for several forecasts, normalised by the climatological RMS variance at each observation point. Dark grey represents displacement error, medium grey is amplitude error, and light grey is residual error. (a) Forecast A2 (2S), (b) C1 (3AV, no wavenumber transition), (c) C2 (3AV, with wavenumber transition), (d) E1 (3SV), (e) E2 (3SV), and (f) G1 (3SV).

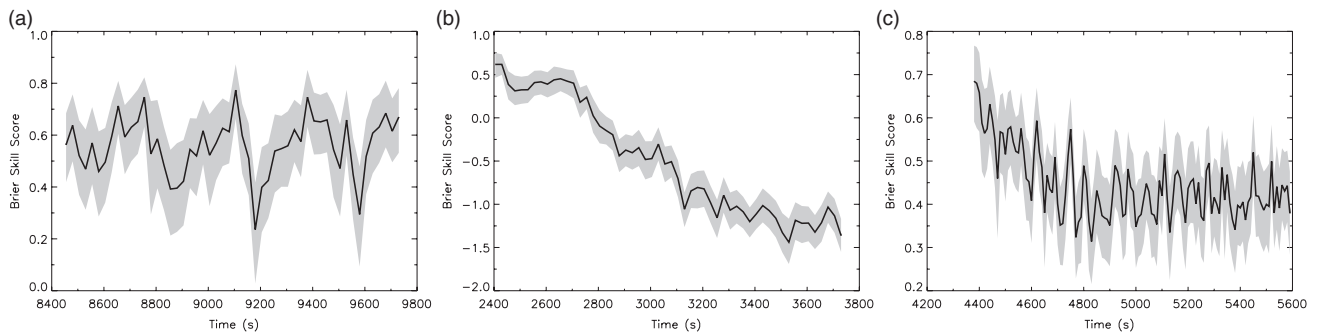


Figure 12. Time series of Brier skill scores for representative forecasts at $z = 9.7$ cm in the three flow regimes. The 95% confidence intervals are shown as shaded regions. (a) forecast A2 (2S), (b) forecast C1 (3AV), and (c) forecast D2 (3SV).

and observations. The normalised RMS error grows to somewhat larger than one but this is not unexpected as the climate of the model and the observations are not expected to be the same.

The decomposition of the RMSE is shown for forecasts C1 and C2 in Figure 11(b,c), showing cases with and without a transition in the control forecast. There is a clear difference in the

character of the decomposition between the two mechanisms of predictability loss. In the forecast drift case (C1), the displacement error becomes dominant quickly due to the drift between forecast and observations. In C2, before the transition at ~ 2800 s there is little displacement error. After the transition, the displaced field still does not offer much of an improvement over the original,

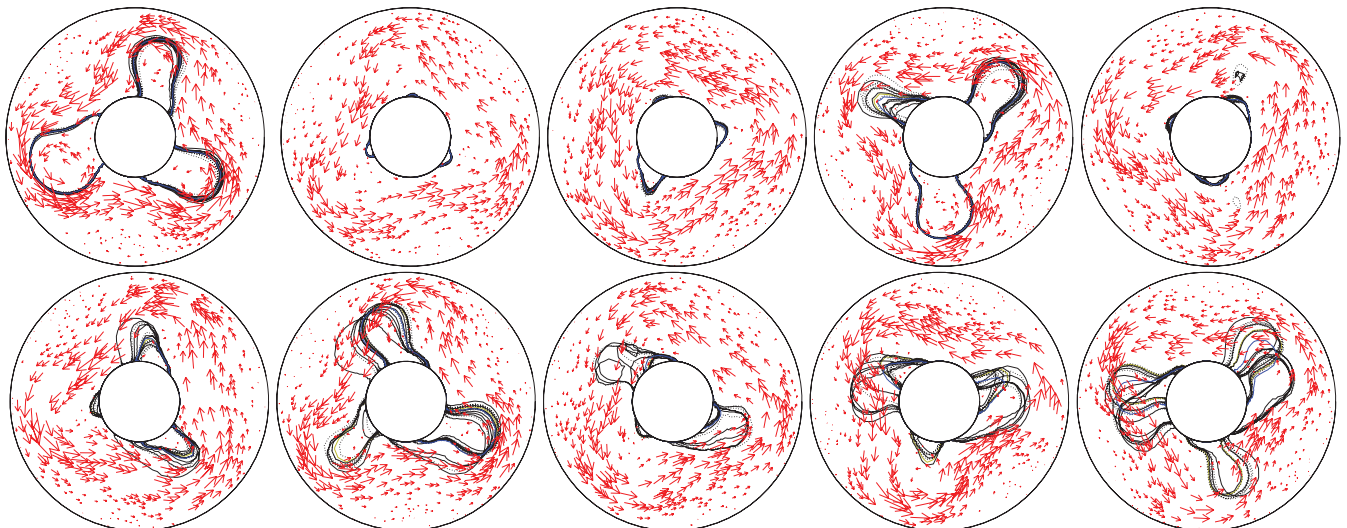


Figure 13. Sequence of spaghetti plots showing forecast B2 (3AV, $\Omega = 0.825 \text{ rad s}^{-1}$). The final frame is at $t_2 + 970$ s. Other details are as Figure 9.

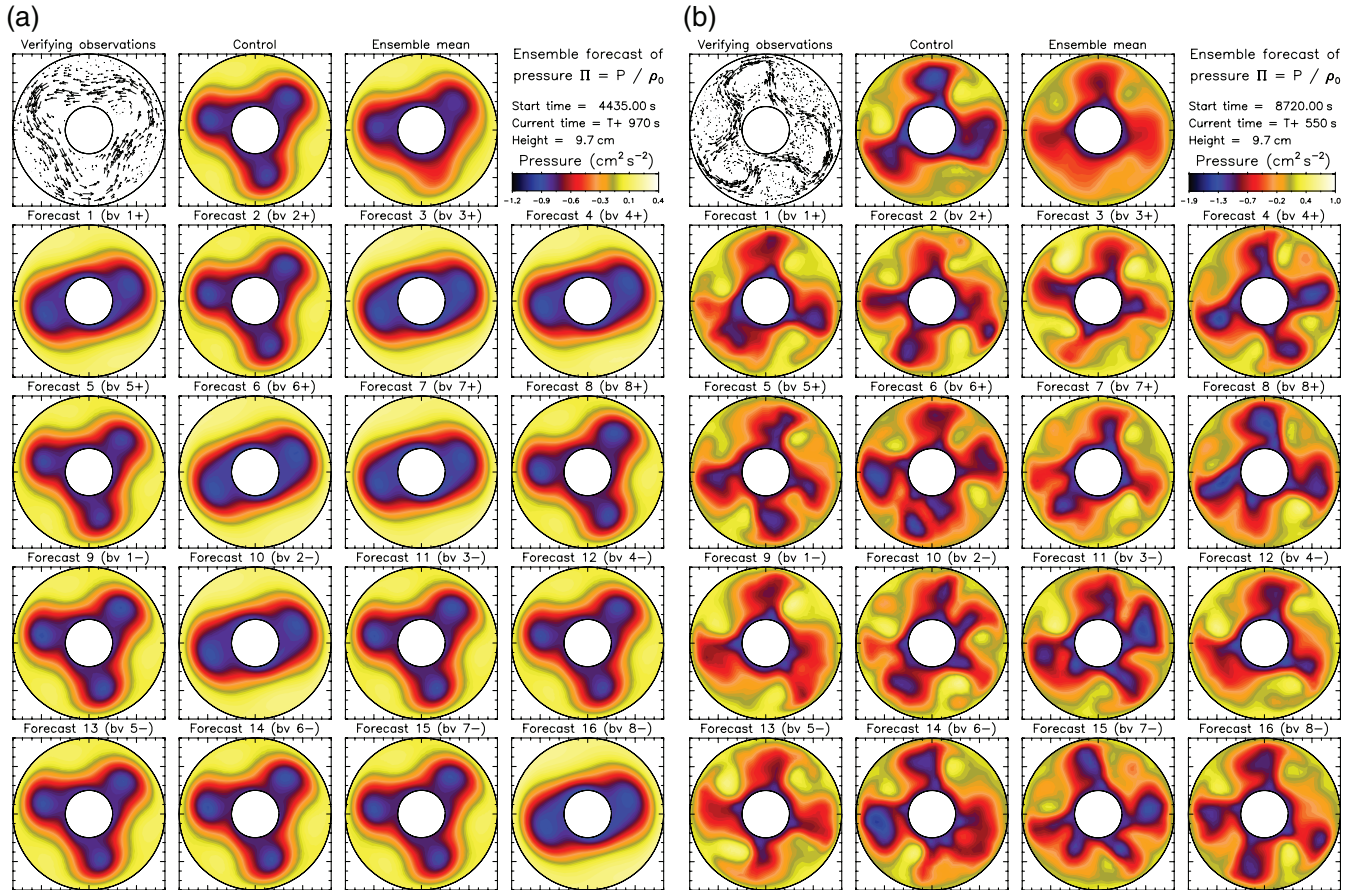


Figure 14. Two ensemble forecasts in the 3AV and 3SV regimes (Figure S10, Document S1, shows other forecasts). Each frame shows the pressure field over a horizontal slice at $z = 9.7$ cm for the control forecast, ensemble mean, eight positive perturbed forecasts, and eight negative perturbed forecasts. For comparison, the observed velocity field is shown in the top left. (a) forecast B2 (3AV) at $t = 5405$ s ($t_2 + 970$ s), and (b) forecast G2 (3SV) at $t = 9270$ s ($t_2 + 550$ s).

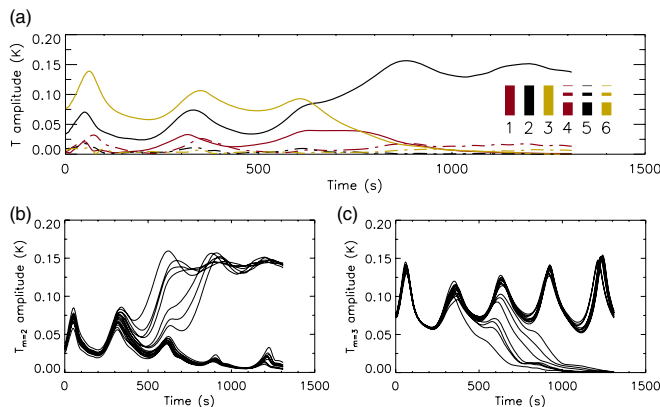


Figure 15. Example of a wavenumber transition in forecast B2. (a) Amplitudes of the first six temperature azimuthal wave modes at mid-height/mid-radius in ensemble member 1+, with colour key. Amplitudes of temperature modes (b) $m = 2$ and (c) $m = 3$ for all ensemble members.

so the displacement error remains small and the residual error becomes large. In C1 the residual error is approximately constant, which is not surprising as, once the shift is corrected for the general shape of the wave, it is well forecast. It was expected *a priori* that amplitude error might be important in this regime as the vacillation is an important feature of the flow, but in general the amplitude error is small compared with the other components.

Figure 12(b) shows the Brier skill score for forecast C1. In general the BSS in this regime is poorer than in the 2S forecasts, with skill being lost over a few hundred seconds. As above, runs which lose predictability by wavenumber transition either have similar (C2 versus C1) or better (B2 versus B1) scores than the forecasts that lose predictability by forecast drift, implying that

wave drift reduces forecast skill more quickly than wavenumber transitions.

3.2.3. Structural vacillation: forecasts D–G

The eight SV forecasts are illustrated by a representative sequence of spaghetti plots in Figure 16, and by part of the G2 forecast in Figure 14(b). It is immediately clear that these forecasts are fundamentally different from the forecasts in the regular regimes. First, the spread of the ensemble is faster than in the regular regimes, which was expected as the model is chaotic in this part of the regime diagram.

Like forecasts B2 and C2, the main feature of these forecasts leading to loss of predictability is wavenumber transition, this time from $m = 3 \rightarrow 4$. Three ensemble members from a typical case are shown in Figure 17, some transitioning and some not. In the 3SV regime there are almost always some ensemble members which transition (except E2, where there are none), which suggests that the model may not be able to sustain these initial conditions. Notice in each case the initial rapid decay of the $m = 3$ mode, even in the case where there is no transition.

Each of the forecasts in this regime follows the same general pattern. First, the dominant wave decays rapidly (over about 200 s) to an almost axisymmetric state; this can be seen in the first frames of Figure 16. The waves then regrow in one of three ways. First, they can reform as $m = 3$ waves, with ensemble members generally in phase (D2, E1, E2, and Figure 16). Second, they can reform as a mixture of $m = 3$ and $m = 4$ waves, with the reformed waves generally in phase (D1, G1, G2). Both these cases imply that the original wave does not decay completely, as otherwise phase information would be lost. Third, the waves reform as a mixture of $m = 3$ and $m = 4$, but with no phase coherence between the reformed waves (F1 and F2).

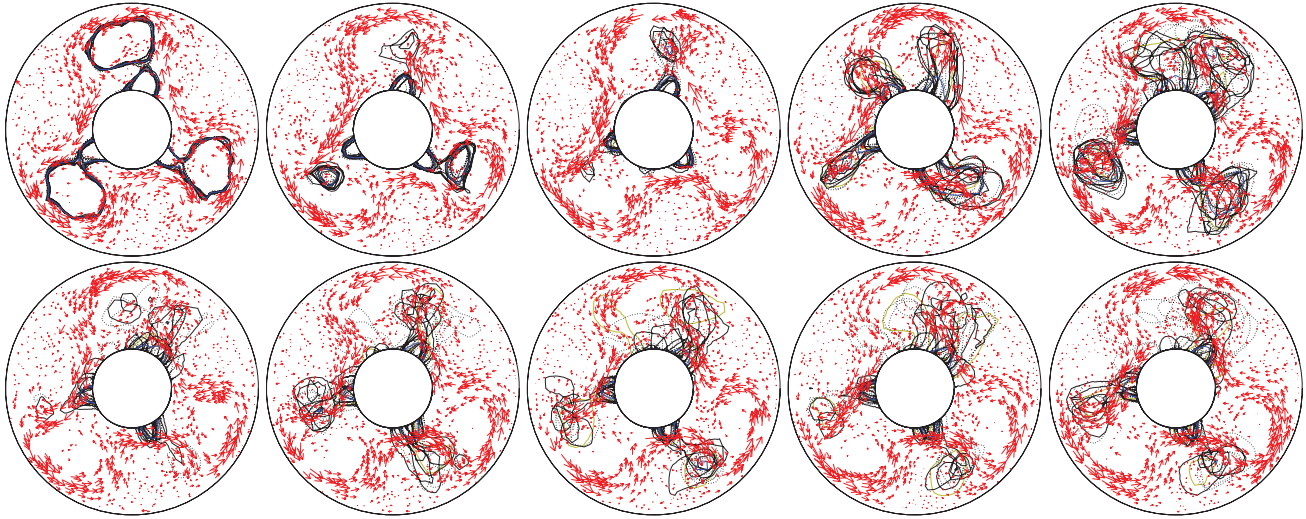


Figure 16. Sequence of spaghetti plots showing forecast E2 (3SV, $\Omega = 2.6 \text{ rad s}^{-1}$). Frames are separated by 60 s, and the final frame is at $t_2 + 550 \text{ s}$. Other details are as Figure 9. Figure S9 of Document S1 shows spaghetti plots for the other forecasts.

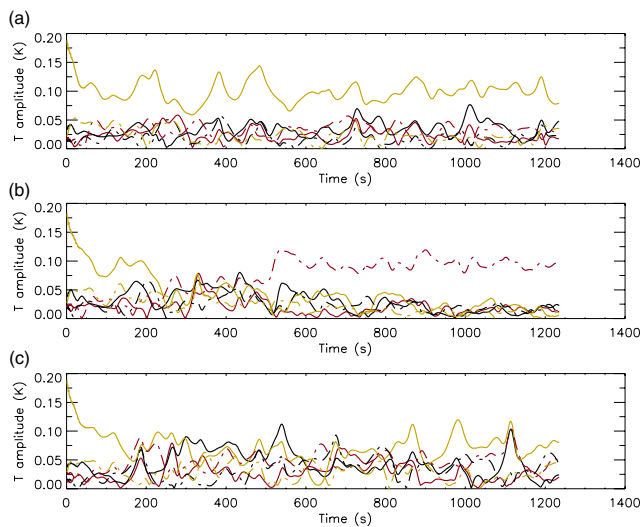


Figure 17. Example of wavenumber transitions in forecast F2, showing the amplitude of the first six azimuthal temperature wave modes over the forecast (mid-height/mid-radius). Colours are the same as Figure 15. (a) Ensemble member 6+ (no transition), (b) ensemble member 8+ (transition to $m = 4$), and (c) ensemble member 4- (final wavenumber unclear).

In each forecast the number of ensemble members that transition from $m = 3 \rightarrow 4$ is variable, from four or fewer in D2, E1, and E2 to twelve or more in D1, F2, and G1. In some forecasts (mainly F and G) there is also some transient $m = 2$ and $m = 5$ behaviour. Sometimes it is not possible to determine the wavenumber at the end of the forecast, e.g. Figure 17(c). The number of ensemble members whose dominant wavenumber cannot be determined generally increases with Ω , and this seems to be the only aspect of the 3SV forecasts that does.

When most of the ensemble members remain $m = 3$ the main source of error is wave drift, but this is generated by a different mechanism to the regular regimes. The drift is caused by a single shift of about 30° in the negative ϕ direction near the start of the forecast as the original waves decay. After this the drift does not change much, because the forecast drift is generally well-estimated for 3SV (Figure 4(c)).

Representative plots of the RMSE are shown in Figure 10(d,e) for forecasts D2 and G2. The main difference from the regular regimes is that there is always a spread of RMSE values in the 3SV forecasts. In each forecast there is an initial rapid increase in RMSE over about 200–300 s before the values plateau for the rest of the forecast; the spread of values reaches a maximum around the same time. The maximum RMSE and the final spread of values is similar for each forecast.

A representative example of the behaviour at a single point is shown in Figure 8(c), typical of all these plots in this regime. The observed values do not change much over the course of the forecast because the wave is phase-locked to the tank (see later). The forecasts remain in the same range as the observations, and the ensemble diverges after a few hundred seconds. Over the 3SV forecasts our general impression is that radial velocity at this point in the flow is reasonably well-estimated, azimuthal velocity is overestimated (too negative), and horizontal velocity magnitude is overestimated (latter two not shown).

The RMSE decomposition takes a number of forms in the 3SV forecasts (Figure 11(d)–(f)), depending on whether the control forecast transitions from $m = 3 \rightarrow 4$ or not (recall that only the control forecast is analysed). In cases when the control forecast transitions (F2 and G1 only, e.g. Figure 11(f)), the displacement error is small compared with the residual error. When the forecast does not transition, the displacement error is either small compared with the residual error (E1 and G2, e.g. Figure 11(d)), or it is initially small but increases rapidly after 100–200 s to 30–50% of the total RMSE (D1, D2, E2, and F1, e.g. Figure 11(e)). Which of these two latter scenarios occurs depends on how the wave is shifted when the dominant wave decays and reforms: if the wave shifts considerably then the displacement error is large.

Figure 12(c) shows the Brier skill score for forecast D2. One might expect that once the wave decays and reforms the forecast ensemble should lose all skill, but this seems not to be the case in the 3SV forecasts. At $z = 9.7 \text{ cm}$ the BSS does not fall to zero within 95% confidence limits, however at $z = 4.3 \text{ cm}$ in several cases $\text{BSS} < 0$ even at the beginning of the forecast, and in all cases skill is generally poorer there.

3.3. Main forecasts: comparative analysis

The RMSE median[†] values for each forecast are shown as time series in Figure 18; there is a clear difference between the three regimes in these plots. The 2S forecasts (black) consistently perform best, and have the smallest spread of values over their ensembles (not shown). The 3AV forecasts (blues) have the poorest scores by the end of the forecasts, although the growth of the RMSE is slower than for 3SV. It is interesting that, in a given flow regime, the RMSE is approximately the same whether

[†]Median is used here instead of the ensemble mean as quite often the ensemble mean performed better than the entire ensemble, as the ensemble mean acts to filter out outliers. The median value here is the median of the individual ensemble members' scores – not the score of the median forecast.

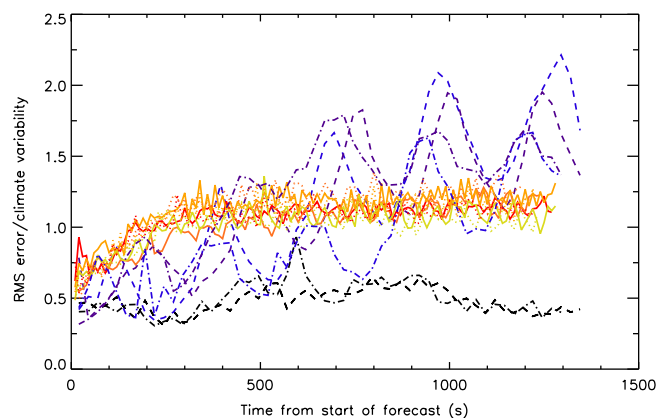


Figure 18. Time series of RMSE for the 14 forecasts, calculated at $z = 9.7$ cm for the horizontal velocity magnitude field from the start of each forecast, and normalised by the climatological RMS variance at each observation point. Colour key is as in Figure 7. Each line represents the median value over the 16 perturbed ensemble members.

predictability is lost via forecast drift or wavenumber transition. The only clear difference between runs B and C is that the partial transition of the ensemble to $m = 2$ in forecast B2 is reflected in the large range of RMSE values in individual ensembles, while in the other cases the range is small. The 3SV forecasts (greens/reds) generally score intermediate between the first two regimes. The initial growth rate of RMSE is faster in 3SV than for 3AV but in the RMSE the saturation value is lower than 3AV. There is little spread of median values between different forecasts, despite the large spread of values within individual ensembles.

These results suggest that the RMSE is robust to small-scale differences between forecasts, and depend more sensitively on the regime the forecast is in. We define our measure of predictability based on forecast *accuracy* by the time taken for the RMSE to reach the climatological RMS variance. At $z = 9.7$ cm this gives typical predictability times of longer than the forecast for 2S, 500 s for 3AV, and 200 s for 3SV.

Time series of the Brier skill score for each of the forecasts are shown in Figure 19. As with the RMSE there is a clear difference between the three regimes. We define the length of a skilful forecast as the time before zero skill is reached within the 95% confidence limits for each forecast (corresponding to BSS around $+0.1$ in each case). Both the 2S and 3SV forecasts retain skill at $z = 9.7$ cm for the whole forecast, with 2S marginally more skilful. The 3AV forecasts lose skill quickly, with BSS going to zero after 300–500 s. Even though the 3SV forecasts diverge rapidly and briefly lose their wave structure, they retain skill throughout the forecasts at the upper observation level. However, at $z = 4.3$ cm even in the 2S regime most forecasts begin with skill either below or just above zero. The spread of skill scores between 3SV forecasts is small at both levels, indicating that these values may represent robust estimates of the general performance of forecasts in this regime, as the values change little with Ω or when the forecasts are repeated.

The difference between the two vertical levels and the skill of the 3SV forecasts can be partially explained using the forecast uncertainty statistic. This is an intrinsic measure of observational uncertainty, is derived from the Brier score (Murphy, 1973; Young, 2010), and is independent of the forecast probabilities. Figure 19 shows this statistic at $z = 9.7$ cm for all the forecasts, with values around 0.15 for the 3SV cases. By contrast, at $z = 4.3$ cm these values are 0.21–0.25, so the 3SV observations are intrinsically more difficult to predict at the lower level. Furthermore, at $z = 9.7$ cm the 3SV uncertainty is lower than for 2S and 3AV, which may help to explain the better BSS scores for 3SV than for 3AV at this level. Based on these results, at $z = 9.7$ cm we found the Brier skill score gives predictability times longer than the forecast for both 2S and 3SV forecasts, and 400 s for 3AV.

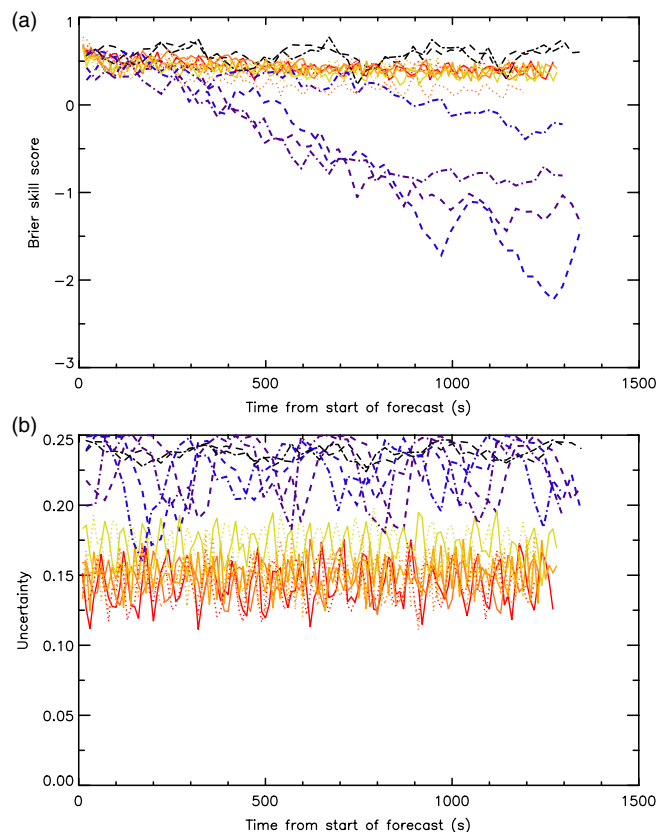


Figure 19. Probability score time series for all 14 forecasts: (a) Brier skill score, and (b) forecast uncertainty obtained by decomposing the Brier score, at $z = 9.7$ cm. Both use the colours in Figure 7.

The three components of the control forecast RMSE decomposition are shown for each forecast in Figure 20. These plots reveal how well the different regimes might be forecast by MORALS if errors from displacement and amplitude were removed. The residual error represents the forecast error once the displacement errors are minimised (by better estimation of the drift rate) and once the amplitude errors are minimised (if future wave vacillation is predicted). Residual error is generally the largest, except where there is significant displacement error. There is little amplitude error except in the 3AV runs.

Using these plots we can try to determine the major sources of error in annulus forecasts. The most interesting plots are of residual error, as they show how the simulation represents the experiment once other errors have been removed. Note that these are typically constant in time, in contrast with the displacement error. To compare the residual error between regimes it is necessary to discard forecasts where the control forecast transitions, because the reason for the residual error is clear in these cases. The remaining cases (Figure 20(d)) show the 2S forecasts have the lowest residual error, followed by the 3AV forecasts, and the 3SV residual error is the largest. These results are different from the scores discussed above, which rank 3AV poorer than 3SV in most cases, but this residual error represents a more robust estimate of the intrinsic error in MORALS' simulations of flow in the rotating annulus. This plot may be the best estimate of the true predictability of the annulus using MORALS, as errors that could be minimised without changing MORALS itself have been removed, and hence the residual errors reflect the model error in the different regimes.

Along with the measures discussed above, three other measures of predictability were computed. The time series of wave amplitudes (e.g. Figures 15 and 17) can be used to compute the time the first wavenumber transition occurs in each ensemble member. The median time over the ensemble is used

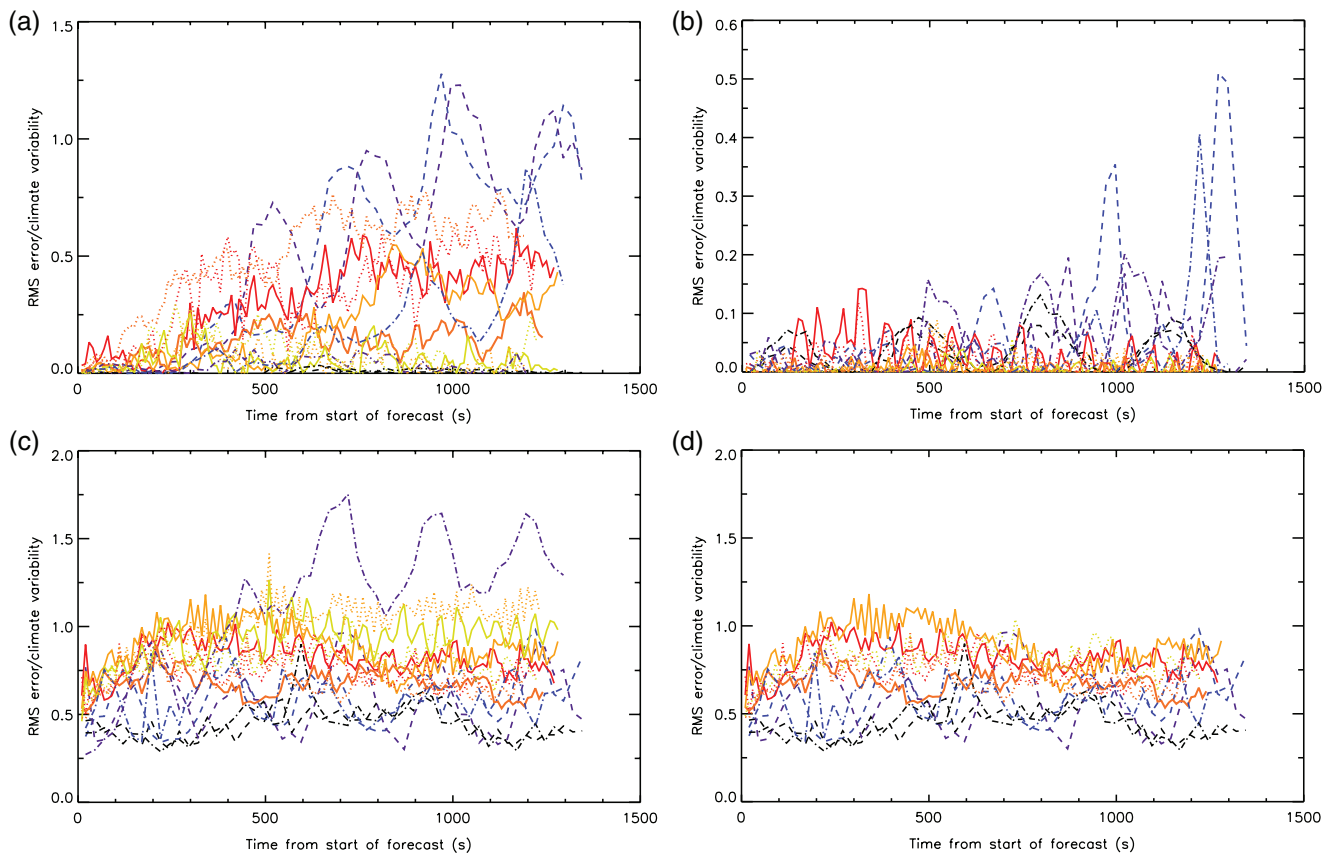


Figure 20. Components of RMSE decomposition for the control run in each of the 14 forecasts. This uses the horizontal velocity magnitude with the colours in Figure 7, and is normalised by the climatological RMS variance at each observation point. (a) Displacement error, (b) amplitude error, (c) residual error, and (d) residual error omitting forecasts whose control runs transition in wavenumber (C2, F2, and G1).

Table 2. List of predictability times (s) for each of the regimes in the FANFARE forecasts, based on the horizontal velocity magnitude $|(u, v)|$ at $z = 9.7$ cm (except single point divergence, at $z = 4.3$ cm). Each estimate is intended to be accurate to within ± 100 s.

Regime	2S	3AV	3SV
Forecast accuracy (RMSE)	>1300	500	200
Forecast skill (BSS)	>1300	400	>1200
Single point divergence time	>1300	500	200
Transition time	N/A	500	200
Wave decay time	N/A	N/A	100

to measure the *transition time*, as it represents the time after which an accurate prediction of the large-scale structure of the flow disappears. Second, the time taken in the 3SV forecasts for the original wave to decay to a minimum can be measured as the *wave decay time*. Finally, we can use the evolution of the ensemble at a single point (specifically the point at $R = 3.0$ cm, $\phi = 0.0$ rad, $z = 4.3$ cm as used in Figure 8), to measure a *single point divergence time*, which is either the time before the ensemble values depart from the observed values, or the time the ensemble begins to diverge. Combining these measures with those above gives a detailed picture of the predictability of the rotating annulus in these flow regimes, which is summarised in Table 2.

4. Wavenumber transitions

A major feature of the 3AV and 3SV forecasts is that in several cases ensemble members transition to other azimuthal wavenumbers. This is an important property of the flow the forecasts failed to predict, so a reason should be sought.

4.1. The shape of the breeding vectors

Some of the breeding vectors have structures that might help to explain why the transitions occur. Note that, if only part of the ensemble transitions, then differences between individual BVs *must* determine which forecasts transition and which do not. In both forecasts B1 and B2, the BVs used to start the forecasts contain perturbations with both $m = 2$ and $m = 3$ structures. A comparison was done in each forecast to see whether there was a correlation between the dominant mode in the BVs and whether the subsequent ensemble member transitioned, but no significant correlation was found.

One case where the BVs did correlate with transitions, however, was forecast B2 (Figure 14(a)). Apart from one pair of \pm perturbations (out of eight), if the positive perturbation transitioned from $m = 3 \rightarrow 2$ then the negative perturbation did not, and vice versa. Forecast B2 was the only 3AV case where the ensemble split in this way.

4.2. A shock to the initial conditions?

A comparison between forecasts in the 3SV regime and a standalone model run using the same parameters suggested that the forecasts are less regular than the standalone run in this regime. When the simulation runs by itself the shape vacillation is less pronounced than at the start of the forecasts, when the flow is disturbed considerably (Figure 16). This suggests some kind of shock occurs in the flow as the forecasts start, as if the simulation cannot support the initial flow state.

This might be a combination of the assimilation method used and the flow regime (it does not happen in the regular regimes). The model is used to update the analysis to produce a new background state between assimilations, but the AC technique does not guarantee the sequence of analyses is consistent with the past evolution of the forecast model. The analysis produced by AC

could be so far off the model manifold in the 3SV regime that the model cannot support the flow structure when left to run by itself.

4.3. The effect of bottom topography

A change in tank geometry could place the forecasts in a different flow regime to the observations. Some regions of the flow in datafiles expf5 and expf6 have no particles at the observation levels – velocities are small there, so particles may sink to the bottom and form small piles, changing the geometry of the annulus and possibly locking the observed wave to the topography. MORALS cannot include bottom topography, so we ran two simulations using the Imperial College Ocean Model (ICOM) (Piggott *et al.*, 2008), with and without bottom topography (Figures S4 and S5, Document S1). The flat-bottomed ICOM simulation did not resemble 3SV flow produced by MORALS at the same point in parameter space, however, or the observations in this regime; the two models seem to have different systematic errors. Therefore it was not possible to make any firm conclusions from these simulations about the effect of topography on the forecasts.

4.4. Hysteresis

Hysteresis is a phenomenon where a system can support more than one regime of behaviour at a particular point in its parameter space, and the behaviour observed depends on how the point in parameter space is approached (Miller and Butler, 1991). It was first observed in annulus experiments in the 1950s (Fultz, 1960).

The AV forecasts are just on the 3AV side of the 2S/3AV transition identified by Young and Read (2008b), but perhaps $m = 3$ is not the only stable wavenumber on the 3AV side of this transition. To test this, we ran a series of simulations keeping ΔT constant at 4.04 K but stepping up Ω by 0.01 rad s⁻¹ every 2000 s, starting at $\Omega = 0.67$ rad s⁻¹ (2S in the Young and Read, 2008b, regime diagram) and finishing at 1.30 rad s⁻¹. A second sequence was run stepping down at the same rate from $\Omega = 0.75$ rad s⁻¹ (3AV in the Young and Read, 2008b, regime diagram) to 0.67 rad s⁻¹. Two thousand seconds appears to be plenty of time for the wave to stabilise at each Ω ; if there is no hysteresis the transition should occur at the same point in both sequences.

Figure 21 shows the results. There is a clear difference in the transition point between stepping-up (Figure 21(a)) and stepping-down (Figure 21(b)). When stepping down, the 3AV \rightarrow 2S transition is at 0.71 rad s⁻¹, but when stepping up 2S is stable until 1.19 rad s⁻¹. So, the simulation can support both $m = 2$ and $m = 3$ at the points in parameter space corresponding to forecasts B and C (0.825 and 0.85 rad s⁻¹). Similar behaviour is often found in experiments (Sitte and Egbers, 2000).

A similar set of simulations were run in the 3SV regime, using $\Delta T = 4.01$ °C. The first series stepped up by 0.2 rad s⁻¹ every 2000 s from 1.6 rad s⁻¹ to 5.0 rad s⁻¹ (Figure 21(c)). The second started at 4.0 rad s⁻¹ and stepped down by 0.2 rad s⁻¹ every 2000 s, ending at 1.6 rad s⁻¹ (Figure 21(d)). These plots also show hysteresis. When the rotation rate is stepped up there is no transition at all: the simulation remains $m = 3$ for the entire sequence, although the wave changes from 3S to 3SV around 2.0 rad s⁻¹. When the rotation rate is stepped down, the simulation begins at 4SV, before transitioning several times to 3SV and back to 4SV before finally changing to 3S at 2.0 rad s⁻¹. The transition point between 3S and 3SV is the same in both cases, but the wavenumber transitions are not. The points in the stepping-down sequence at the rotation rates used in the forecasts (2.4, 2.6, 2.8, and 3.0 rad s⁻¹) are all predominantly 4SV, while in the stepping-up sequence they are 3SV, and therefore the simulation can support both waves at the forecast parameters.

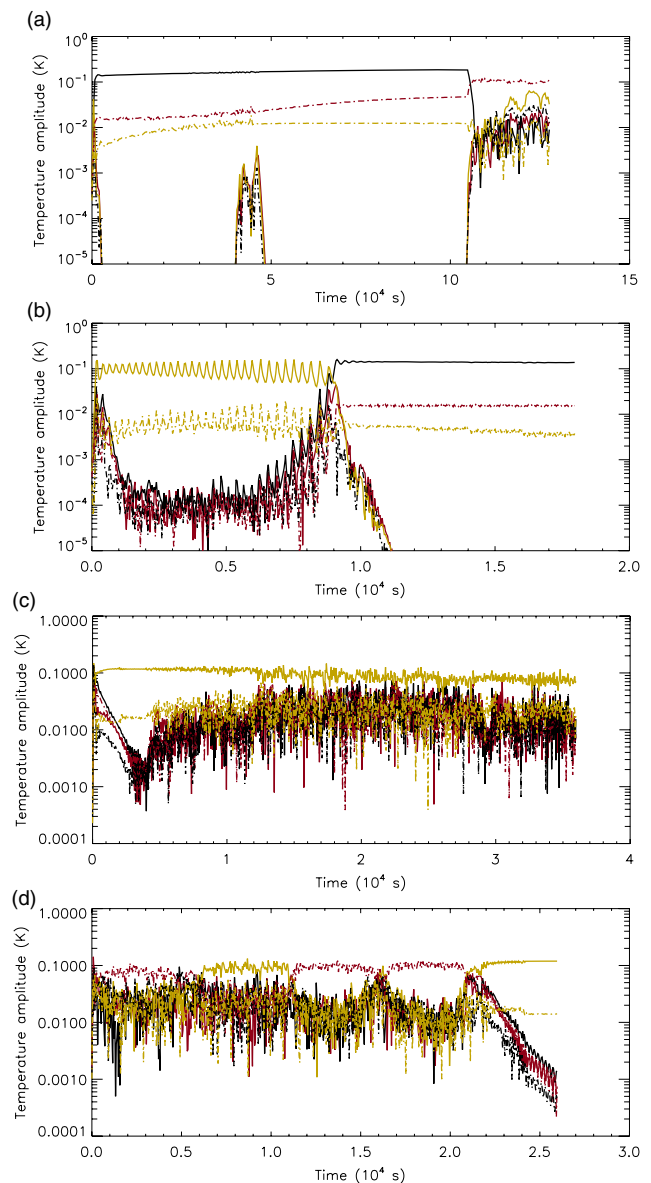


Figure 21. Amplitudes of the first six azimuthal temperature modes at mid-height/mid-radius during hysteresis runs. Colours are the same as Figure 15. (a) Stepping up: 2S \rightarrow 4SV transition (0.67 \rightarrow 1.30 rad s⁻¹, testing 2S/3AV boundary), (b) stepping down: 3AV \rightarrow 2S transition (0.75 \rightarrow 0.67 rad s⁻¹, testing 2S/3AV boundary), (c) stepping up: high Ω (1.6 \rightarrow 5.0 rad s⁻¹, testing 3SV/4SV boundary), and (d) stepping down: high Ω (4.0 \rightarrow 1.6 rad s⁻¹, testing 3SV/4SV boundary).

4.5. Summary

We have gone some way towards explaining why transitions occur in the forecasts, but no definitive reason has been found. Taken together, perhaps something like Figure 22 is happening. The hysteresis tests show that, at the relevant points in the MORALS parameter space, there exist multiple model attractors. On the model manifold there may be regions of attraction, represented in Figure 22 by minima in a potential surface, and each simulation will eventually fall into one of these attractors depending on its initial condition. A model trajectory will move towards the model manifold over time, but an analysis produced using analysis correction will almost surely not lie on it initially. Unavoidable model error means that the possible states admitted by the model (the model climatology) and the possible states of the system (the ‘true’ climatology) are disjoint; the analysis pulls the state away from the model manifold towards the ‘true’ manifold, with the result that the analysis lies on neither. If the position of the analysis relative to the model manifold moves the initial condition out of its original basin of attraction, or if the perturbation does so (as shown in the figure), then the subsequent forecast can tend

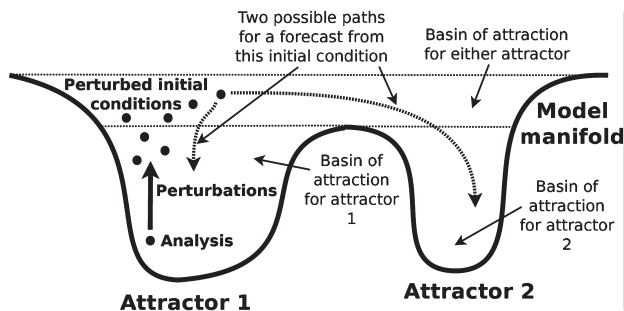


Figure 22. Schematic of a possible model manifold explanation for the wavenumber transitions in the forecasts.

towards either attractor over time, leading to a possible regime transition.

Finally, because the forecasts are probabilistic, the cases where there are regime transitions may in fact reflect the *a priori* probability of the real experiments ending up in one of the two (or more) basins of attraction. Several experimental regime diagrams indicate that more than one regime can be realised at certain points in parameter space (Hignett *et al.*, 1985), so the ensemble forecasts may therefore be more reliable than at first glance.

5. Conclusions

In this article we addressed two questions about the rotating annulus laboratory experiment: how predictable is this system, and can common meteorological techniques be used to study its behaviour? Building on work by Young Read (2008a, 2008b, 2013) the annulus simulation MORALS, ensemble generation using breeding vectors, and data assimilation using analysis correction were combined into a full framework for ensemble forecasting in the laboratory annulus. This framework was used to predict the behaviour of regular (2S, 3AV) and chaotic (3SV) rotating annulus flow, verifying forecasts against laboratory data.

In general, the relative predictability of the flow regimes was clear from these forecasts. The 2S forecasts performed well, and in both 3AV and 3SV forecasts quality and skill was lost either by azimuthal drift of the dominant wave mode or by a transition away from the dominant wavenumber in the observations. Several hypotheses were proposed to explain the transitions, but no main underlying cause was identified. The model displays hysteresis in the relevant regions of parameter space, so one possibility is that these forecasts might reflect an underlying distribution of the regimes that can be realised at that point in parameter space. Given sufficient experimental time, this could be tested in the laboratory.

The flow was typically predictable up to the end of the forecasts in the 2S regime, for 400–500 s in the 3AV regime, and for 100–300 s in the 3SV regime (Table 2). It is useful to compare these numbers with other, arguably more ‘intrinsic’, measures of predictability. Young and Read (2008a) estimated annulus Lyapunov exponents in the PMS, and they found the Lyapunov time to be about 4000 and 300 s in regular and chaotic regimes respectively. Second, regular regime BVs have a doubling time around 200 s, and the chaotic regime BVs around 70 s (Figure 7). The relative values of these measures between regular and chaotic regimes are reflected in our forecast results.

There are places where the forecasts could be improved. The drift correction significantly improved the quality of the forecasts, but it could be made more accurate, as the drift rate varies considerably between regimes (Figure 4). For example, accounting for the vacillation cycle by using a sinusoidal function $A + Bt + C \sin(t/D + E)$ for the drift instead of the current linear fit should improve the forecasts in that regime significantly. Also some form of initialization is usually common in data assimilation

(Charney, 1955; Phillips, 1960), to suppress imbalances between the fields. AC uses the nudging technique (Hoke and Anthes, 1976) as an implicit initialization method, but others could be tried. Finally, ensemble generation based on parameter perturbation could be attempted. This would perturb the values of Ω and ΔT used for the forecasts instead of the initial conditions. Such an approach might be useful because the Ω shift between observed and simulated parameter space is only known to within about $\pm 0.02 \text{ rad s}^{-1}$.

This work has shown that methods used for forecasting planetary atmospheres through operational weather forecasting can also be applied to the laboratory experiment. These methods were set up in a similar way to their operational implementation and were used to forecast data from the laboratory experiment. The next stage would be to identify specific new methods that might benefit from rigorous analysis in this system. Perhaps the most suitable would be those that are not yet in operational use, because those that are have already been tested under many situations encountered in operational practice. One such promising technique is gradient descent/shadowing (Judd, 2003; Judd *et al.*, 2008).

One other method that is now popular is the Ensemble Kalman Filter (EnKF; Evensen, 1994). Zhang and Snyder (2007) list a number of challenges which implementation of the EnKF had not overcome at the time of writing, and some of these could be addressed using the annulus. Model error and bias are major challenges for the EnKF because the model is used to propagate the error covariances forward in time. It might be fruitful to study this in the annulus, where specific sources of model error might be more easily identified than in atmospheric models. The EnKF has already been implemented in the annulus by Ravela *et al.* (2010), so their method might be a good place to start.

Acknowledgements

This work was financially supported by NERC Studentship NER/S/A/2005/13667. It describes the main results of RMBY's DPhil thesis and there is a full list of acknowledgments there. In particular, RMBY thanks David Andrews and Stephen Lewis for examining the thesis, João Mendonça for computing resources, and James Maddison for running the ICOM simulations. Many figures used David Fanning's Coyote IDL Program Library. We thank two anonymous reviewers for their comments.

Supporting information

The following supporting information is available as part of the online article:

Document S1. Additional information covering some technical points on datafile expf1, climate distributions, parameter testing, ICOM simulations, flow diagrams of the different forecast stages, and complete sets of figures for each forecast.

References

- Abbe C. 1907. Projections of the globe appropriate for laboratory methods of studying the general circulation of the atmosphere. *Bull. Am. Math. Soc.* **13**: 502–506, doi: 10.1090/S0002-9904-1907-01512-4.
- Bigelow FH. 1902. Studies on the statics and kinematics of the atmosphere in the United States: IV.-Review of Ferrel's and Oberbeck's theories of the local and the general circulations. *Mon. Weather Rev.* **30**: 163–171, doi: 10.1175/1520-0493(1902)30[163:SOTSAK]2.0.CO;2.
- Brier GW. 1950. Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* **78**: 1–3, doi: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.
- Buizza R, Palmer TN. 1995. The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.* **52**: 1434–1456, doi: 10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2.
- Charney J. 1955. The use of the primitive equations of motion in numerical prediction. *Tellus* **7**: 22–26, doi: 10.1111/j.2153-3490.1955.tb01138.x.

- Corazza M, Kalnay E, Patil DJ, Yang SC, Morss R, Cai M, Szunyogh I, Hunt BR, Yorke JA. 2003. Use of the breeding technique to estimate the structure of the analysis 'errors of the day'. *Nonlinear Process. Geophys.* **10**: 233–243, doi: 10.5194/npg-10-233-2003.
- Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**: 10143–10162, doi: 10.1029/94JC00572.
- Farnell L, Plumb RA. 1976. *Numerical Integration of Flow in a Rotating Annulus II: Three-Dimensional Model*. Met O 21 Occasional Note 76/1. Geophysical Fluid Dynamics Laboratory, Meteorological Office: Bracknell, UK.
- Früh WG, Read PL. 1997. Wave interactions and the transition to chaos of baroclinic waves in a thermally driven rotating annulus. *Phil. Trans. R. Soc. London, Ser. A* **355**: 101–153, doi: 10.1098/rsta.1997.0003.
- Fultz D. 1951. Non-dimensional equations and modeling criteria for the atmosphere. *J. Meteorol.* **8**: 262–267, doi: 10.1175/1520-0469(1951)008<0262:NDEAMC>2.0.CO;2.
- Fultz D. 1960. Experimental models of rotating fluids and possible avenues for future research. In *Dynamics of Climate*, Pfeffer RL. (ed.): 71–77. Pergamon Press: Oxford, UK, doi: 10.1016/B978-1-4831-9890-3.50019-0.
- Gilmour I. 1998. 'Nonlinear model evaluation: *t*-shadowing, probabilistic prediction and weather forecasting', DPhil thesis. University of Oxford: UK.
- Gilmour I, Smith LA, Buizza R. 2001. Linear regime duration: Is 24 hours a long time in synoptic weather forecasting? *J. Atmos. Sci.* **58**: 3525–3539, doi: 10.1175/1520-0469(2001)058<3525:LRDIHA>2.0.CO;2.
- Harlim J, Oczkowski M, Yorke JA, Kalnay E, Hunt BR. 2005. Convex error growth patterns in a global weather model. *Phys. Rev. Lett.* **94**: 228501, doi: 10.1103/PhysRevLett.94.228501.
- Hide R. 1953. Some experiments on thermal convection in a rotating liquid. *Q. J. R. Meteorol. Soc.* **79**: 161–161, doi: 10.1002/qj.4970793916.
- Hide R, Mason PJ. 1975. Sloping convection in a rotating fluid. *Adv. Phys.* **24**: 47–100, doi: 10.1080/00018737500101371.
- Hignett P, White AA, Carter RD, Jackson WDN, Small RM. 1985. A comparison of laboratory measurements and numerical simulations of baroclinic wave flows in a rotating cylindrical annulus. *Q. J. R. Meteorol. Soc.* **111**: 131–154, doi: 10.1002/qj.49711146705.
- Hoffman RN, Liu Z, Louis JF, Grassotti C. 1995. Distortion representation of forecast errors. *Mon. Weather Rev.* **123**: 2758–2770, doi: 10.1175/1520-0493(1995)123<2758:DROFE>2.0.CO;2.
- 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.
- Houtekamer PL, Derome J. 1994. Prediction experiments with two-member ensembles. *Mon. Weather Rev.* **122**: 2179–2191, doi: 10.1175/1520-0493(1994)122<2179:PEWTME>2.0.CO;2.
- Jacoby TNL, Read PL, Williams PD, Young RMB. 2011. Generation of inertia-gravity waves in the rotating thermal annulus by a localised boundary layer instability. *Geophys. Astrophys. Fluid Dyn.* **105**: 161–181, doi: 10.1080/03091929.2011.560151.
- Judd K. 2003. Nonlinear state estimation, indistinguishable states, and the extended Kalman filter. *Physica D* **183**: 273–281, doi: 10.1016/S0167-2789(03)00180-5.
- Judd K, Reynolds CA, Rosmond TE, Smith LA. 2008. The geometry of model error. *J. Atmos. Sci.* **65**: 1749–1772, doi: 10.1175/2007JAS2327.1.
- Lorenz 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.
- Lorenz EN. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**: 130–141, doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.
- Lorenz EN. 1975. Climatic predictability. In *The Physical Basis of Climate and Climate Modelling*, GARP Publication Series 16: 132–136. World Meteorological Organisation: Geneva, Switzerland.
- Mason SJ. 2008. Understanding forecast verification statistics. *Meteorol. Appl.* **15**: 31–40, doi: 10.1002/met.51.
- Miller TL, Butler KA. 1991. Hysteresis and the transition between axisymmetric flow and wave flow in the baroclinic annulus. *J. Atmos. Sci.* **48**: 811–823, doi: 10.1175/1520-0469(1991)048<0811:HATTBA>2.0.CO;2.
- 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.
- Murphy AH. 1973. A new vector partition of the probability score. *J. Appl. Meteorol.* **12**: 595–600, doi: 10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2.
- Newman CE, Read PL, Lewis SR. 2004. Investigating atmospheric predictability on Mars using breeding vectors in a general-circulation model. *Q. J. R. Meteorol. Soc.* **130**: 2971–2989, doi: 10.1256/qj.03.209.
- Phillips NA. 1960. On the problem of initial data for the primitive equations. *Tellus* **12**: 121–126, doi: 10.1111/j.2153-3490.1960.tb01289.x.
- Piggott MD, Gorman GJ, Pain CC, Allison PA, Candy AS, Martin BT, Wells MR. 2008. A new computational framework for multi-scale ocean modelling based on adapting unstructured meshes. *Int. J. Numer. Methods Fluids* **56**: 1003–1015, doi: 10.1002/fld.1663.
- Ravela S, Marshall J, Hill C, Wong A, Stransky S. 2010. A realtime observatory for laboratory simulation of planetary flows. *Exp. Fluids* **48**: 915–925, doi: 10.1007/s00348-009-0752-0.
- Read PL, Bell MJ, Johnson DW, Small RM. 1992. Quasi-periodic and chaotic flow regimes in a thermally driven, rotating fluid annulus. *J. Fluid Mech.* **238**: 599–632, doi: 10.1017/S0022112092001836.
- Read PL, Thomas NPJ, Risch SH. 2000. An evaluation of Eulerian and semi-Lagrangian advection schemes in simulations of rotating, stratified flows in the laboratory. Part I: Axisymmetric flow. *Mon. Weather Rev.* **128**: 2835–2852, doi: 10.1175/1520-0493(2000)128<2835:AE0EAS>2.0.CO;2.
- Read PL, Pérez EP, Moroz IM, Young RMB. 2015. General circulation of planetary atmospheres: Insights from rotating annulus and related experiments. In *Modeling Atmospheric and Oceanic Flows*, von Larcher T, Williams PD. (eds.): 9–44, doi: 10.1002/9781118856024.ch1. John Wiley & Sons: Hoboken, NJ.
- Rogberg P, Read PL, Lewis SR, Montabone L. 2010. Assessing atmospheric predictability on Mars using numerical weather prediction and data assimilation. *Q. J. R. Meteorol. Soc.* **136**: 1614–1635, doi: 10.1002/qj.677.
- Sitte B, Egbers C. 2000. Higher order dynamics of baroclinic waves. *Lect. Notes Phys.* **549**: 355–375, doi: 10.1007/3-540-45549-3_20.
- Thomson J. 1892. Bakerian lecture: On the grand currents of atmospheric circulation. *Philosophical Transactions of the Royal Society of London A* **183**: 653–684, doi: 10.1098/rsta.1892.0017.
- Toth Z, Kalnay E. 1993. Ensemble forecasting at NMC: The generation of perturbations. *Bull. Am. Meteorol. Soc.* **74**: 2317–2330, doi: 10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.
- Toth Z, Kalnay E. 1997. Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.* **125**: 3297–3319, doi: 10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2.
- Vallis GK. 2006. *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*. Cambridge University Press: Cambridge, UK.
- Vettin F. 1857. Ueber den aufsteigenden Luftstrom, die Entstehung des Hagels und der Wirbel-Stürme. *Ann. Phys.* **178**: 246–255, doi: 10.1002/andp.18571781005.
- Weigel AP, Liniger MA, Appenzeller C. 2007. The discrete Brier and ranked probability skill scores. *Mon. Weather Rev.* **135**: 118–124, doi: 10.1175/MWR3280.1.
- Wilks DS. 2001. A skill score based on economic value for probability forecasts. *Meteorol. Appl.* **8**: 209–219, doi: 10.1017/S1350482701002092.
- Woodcock F. 1976. The evaluation of yes/no forecasts for scientific and administrative purposes. *Mon. Weather Rev.* **104**: 1209–1214, doi: 10.1175/1520-0493(1976)104<1209:TEOYFF>2.0.CO;2.
- Young RMB. 2010. Decomposition of the Brier score for weighted forecast-verification pairs. *Q. J. R. Meteorol. Soc.* **136**: 1364–1370, doi: 10.1002/qj.641.
- Young RMB, Read PL. 2008a. Breeding and predictability in the baroclinic rotating annulus using a perfect model. *Nonlinear Process. Geophys.* **15**: 469–487, doi: 10.5194/npg-15-469-2008.
- Young RMB, Read PL. 2008b. Flow transitions resembling bifurcations of the logistic map in simulations of the baroclinic rotating annulus. *Physica D* **237**: 2251–2262, doi: 10.1016/j.physd.2008.02.014.
- Young RMB, Read PL. 2013. Data assimilation in the laboratory using a rotating annulus experiment. *Q. J. R. Meteorol. Soc.* **139**: 1488–1504, doi: 10.1002/qj.2061.
- Young RMB, Read PL, Früh WG, Smith D, Risch SH. 2015. *The thermally-driven rotating annulus: horizontal velocities in regular and weakly chaotic flow regimes [Dataset]*. Oxford University Research Archive: Oxford, UK, doi: 10.5287/bodleian:dr26xx49n.
- Zhang F, Snyder C. 2007. Ensemble-based data assimilation. *Bull. Am. Meteorol. Soc.* **88**: 565–568, doi: 10.1175/BAMS-88-4-565.