**Nonlinear Processes
in Geophysics**

Breeding and predictability in the baroclinic rotating annulus using a perfect model

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

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

Abstract. We present results from a computational study of predictability in fully-developed baroclinically unstable laboratory flows. This behaviour is studied in the Met Office/Oxford Rotating Annulus Laboratory Simulation – a model of the classic rotating annulus laboratory experiment with differentially heated cylindrical sidewalls, which is firmly established as an insightful laboratory analogue for certain kinds of atmospheric dynamical behaviour. This work is the first study of "predictability of the first kind" in the annulus experiment. We devise an ensemble prediction scheme using the breeding method to study the predictability of the annulus in the perfect model scenario. This scenario allows one simulation to be defined as the true state, against which all forecasts are measured. We present results from forecasts over a range of quasi-periodic and chaotic annulus flow regimes. A number of statistical and meteorological techniques are used to compare the predictability of these flows: bred vector growth rate and dimension, error variance, "spaghetti plots", probability forecasts, Brier score, and the Kolmogorov-Smirnov test. These techniques gauge both the predictability of the flow and the performance of the ensemble relative to a forecast using a climatological distribution. It is found that in the perfect model scenario, the two quasi-periodic regimes examined may be indefinitely predictable. The two chaotic regimes (structural vacillation and period doubled amplitude vacillation) show a loss of predictability on a timescale of hundreds to thousands of seconds (65–280 annulus rotation periods, or 1–3 Lyapunov times).

1 Introduction

Knowledge of short to medium-term atmospheric predictability is important for meteorologists and society in general. How far ahead can we predict the weather with accuracy? Quantification is difficult and complex in real atmospheres, so we need simpler analogues for testing ideas and methods. The rotating annulus (Hide, 1953; Hide and Mason, 1975; Read et al., 1992, and Fig. 1) is a laboratory system which has been used for some 50 years to produce behaviour qualitatively similar to the mid-latitudes of a generic planetary atmosphere. It is a useful test bed for certain kinds of atmospheric dynamical behaviour and the methods used to study them, and displays interesting behaviour worthy of study in its own right. This paper details the first stages of an investigation into the predictability of baroclinically unstable flow in this system. Baroclinic instability is important in the Earth's midlatitudes for the large scale transport of heat and momentum. It can be reproduced in the annulus under certain conditions, and the laboratory setting allows this to be studied in a controlled and reproducible manner. The main aim of this work is to investigate the breakdown in predictability of these flows in the laboratory in a range of quasi-periodic and chaotic regimes.

Our approach to predictability is to forecast annulus behaviour using techniques common to weather forecasting – an ensemble prediction scheme, combined with a numerical forecast model of the annulus itself. Lorenz (1975) defined two kinds of predictability: the first kind concerns the future evolution of a system from initial conditions, and the second kind concerns predicted behaviour given certain boundary conditions (i.e. prediction of the climate or attractor of the system, or in the annulus context, prediction of the flow regime given a particular experimental setup). This second kind of predictability 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; Young

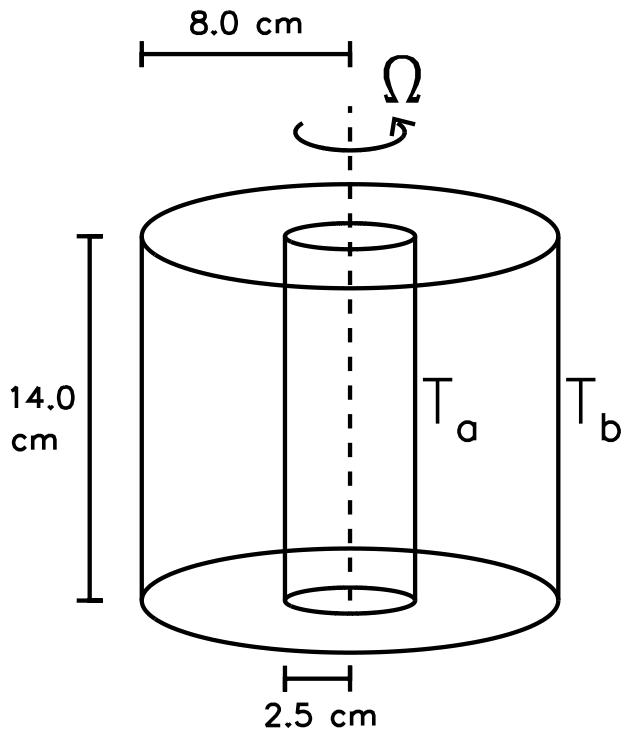*Correspondence to:* R. M. B. Young
(young@atm.ox.ac.uk)

Fig. 1. Schematic (to scale) showing the "standard" rotating annulus setup, with inner and outer cylinders at temperatures T_a and T_b respectively, rotating with constant angular velocity Ω. Fluid is contained between the cylinders.

and Read, 2008, for example). Prediction of the first kind, however, has not previously been studied in any depth for the annulus system, and we believe this study is the first to do so. Some previous studies have calculated the Lyapunov exponents for various annulus flow regimes using experimental time series (Read et al., 1992; Früh and Read, 1997) and simulated time series (Young and Read, 2008). The Lyapunov exponent is a measure of first-kind predictability, but these exponents have only limited usefulness for characterizing the general predictability of complex systems.

This paper reports exclusively on computational work. We devise an ensemble prediction scheme based on the breeding method (Toth and Kalnay, 1993, 1996, 1997) to explore the predictability of the annulus in the perfect model scenario. In this scenario, we assume that our model is a perfect representation of the annulus experiment; equivalently, the "data" used to verify our forecasts and to set up our ensemble are generated using the same simulation as the forecasts themselves, and are assumed to be exact. No model is truly perfect, but with care this scenario may be applied as long as its limitations are appreciated. In this context, the main advantages of the perfect model scenario are that forecast errors can be calculated explicitly, and that those errors are due only to the uncertainty in the initial conditions and the dynamics of the flow under study. Practical complexities associated with including a data assimilation scheme are avoided because the "true" state of the system is known, so data assimilation is not required. The broader goal for follow-up work is to extend this scheme into the more realistic imperfect model scenario, which will requires assimilating laboratory data into the forecasts. The work in this paper is a necessary step towards that goal, and assuming a perfect model is a common technique for working under simplified conditions (Houtekamer and Derome, 1994, for example). Using a similar scheme with real data and an imperfect model will then allow the practical predictability of the laboratory system to be estimated, and hindcasts to be made of the annulus flow.

The method for choosing ensemble initial conditions is important in numerical weather prediction (NWP). The distribution of perturbations about the analysis should best represent the analysis errors, so that the distribution of potential states is well-sampled by the ensemble. In this work we use the breeding method to generate the ensemble. Breeding was developed as a computationally inexpensive way to perturb the atmospheric analysis and to follow those perturbations in a way reflecting the previous non-linear evolution of the flow. The resulting error fields (the "breeding vectors") identify regions of most rapidly growing instability. Its first use was in the National Meteorological Center operational ensemble (Tracton and Kalnay, 1993), and has been used in a number of other contexts since then (see Sect. 3).

We will examine predictability using a range of qualitative and quantitative measures: evolution of the forecast error variance gives a simple measurement of the ensemble spread; the bred vector growth rate is related to the leading Lyapunov exponent (Kalnay et al., 2002), and can be used as a measure of predictability in a similar way; the bred vector dimension (Patil et al., 2001) has been shown to be directly related to predictability in simple systems, and we shall see whether this is also the case in the annulus model; "spaghetti plots" (Tracton and Kalnay, 1993) allow the forecast evolution to be visualized clearly; Kolmogorov-Smirnov goodness of fit statistics (Massey, 1951) compare the evolution of an ensemble of forecasts with a climatological forecast, so can be used to estimate when the "usefulness" of the forecast is exhausted; probability forecasts visualize how predictive power varies in space, and the Brier score (Brier, 1950) verifies these probabilities against the actual outcome. Neither the absolute nor the relative predictability of a system should depend on the method used to measure it. Therefore, we expect each of these measures to give the same conclusions, certainly about the relative predictability of different flow regimes. If they do not give the same conclusions then their usefulness as a measure of predictability may be questionable.

A secondary aim of this work concerns NWP. Few attempts have taken advantage of the properties of laboratory systems mentioned above to inform the development of operational forecasting techniques, and this study is the first to

apply NWP methodology in the controlled and reproducible environment provided by the annulus. We believe that because of the properties of the rotating annulus, it could provide useful insights in the future for improving existing approaches to assimilation and forecasting, by investigating and comparing techniques in current use for NWP, but in a system where much of the behaviour can be controlled. An additional advantage of the annulus setup is that non-chaotic flow can be generated. Operational NWP is almost exclusively restricted to chaotic flow, so the techniques are little-used in non-chaotic systems, and doing so might provide some previously-unseen insights into their use.

We are aware that using the annulus as a direct analogue for the Earth's highly turbulent atmosphere is questionable, but as the annulus can represent a wide range of rotational and temperature forcing, the analogue may be closer to the situation on other planets such as Mars (where NWP techniques are also applicable, of course). Observations and numerical simulations of the Martian atmosphere (see Leovy, 1985, for a review) have generally concluded that baroclinic waves are much more regular on Mars than on Earth (Barnes, 1981; Collins and James, 1995; Collins et al., 1996). Baroclinic flow on Mars may therefore be better represented by the rotating annulus analogue than baroclinic flow on Earth (Read et al., 1998). Recent work by Newman et al. (2004) also concluded that the Martian atmosphere is at times more predictable than the Earth's – is there something fundamentally different about the Martian atmosphere which determines this behaviour? Comparing annulus flow regimes similar to the atmospheres of the Earth and Mars may also allow aspects of this question to be tackled.

Section 2 briefly describes the simulation used to model the rotating annulus. Section 3 outlines the ensemble prediction scheme developed to forecast annulus flow in the perfect model scenario. The results from ten ensemble forecasts in a range of different regimes are presented in Sect. 4, and our findings are discussed and concluded in Sect. 5.

2 The simulation

We use the Met Office/Oxford Rotating Annulus Laboratory Simulation (MORALS) code (Farnell and Plumb, 1976; Hignett et al., 1985; Read et al., 2000)[1], which is well-established as a comprehensive and quantitatively accurate model for dynamical behaviour in the rotating annulus, at least in regular and weakly vacillating regimes (Read et al., 1997). It solves the Navier-Stokes, mass continuity, and heat transfer equations along with a Poisson equation for pressure, and equations of state for density, viscosity and thermal diffusivity (Table 1), under the Boussinesq approxima-

---

[1] Farnell, L. and Plumb, R. A.: Numerical integration of flow in a rotating annulus II: three dimensional model, Occasional Note Met O 21 76/1, Meteorological Office, Bracknell, UK, unpublished, 24 pp., 1976.
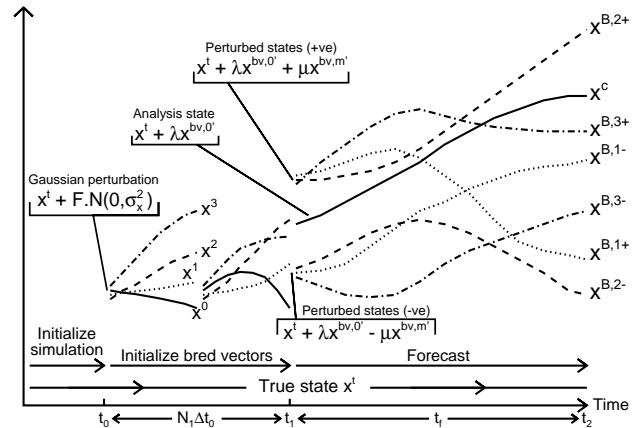

Fig. 2. A schematic of the ensemble prediction scheme, showing the evolution of each component of the forecast as a function of time during the three stages described in the text. The bred vectors are initialised in stage one by perturbing the truth x^t. The bred vectors are then added to the truth in stage two to form the control forecast x^c and the perturbed forecasts $x^{B,m\pm}$. These forecasts are then verified against the truth during the forecast stage.

tion for a cylindrical fluid annulus rotating at angular velocity Ω. The equations are cast in velocity-temperature-pressure form: u (radial), v (azimuthal) and w (vertical) velocities/cm s^{-1}, temperature $T/$°C and a scaled pressure $\Pi = P/\rho_0/$cm^2 s^{-2}. T is defined relative to a reference temperature T_0, and Π is relative to a reference pressure $\Pi_0(R, z) = \frac{1}{2}\Omega^2 R^2 + g(d - z)$. We use the "standard" configuration (Fig. 1) with inner and outer walls maintained at constant temperatures T_a and T_b respectively, with no internal heating, and include a 'rigid lid' (a no-slip upper boundary condition). The model grid is defined using cylindrical polar coordinates (R, ϕ, z), stretched in R and z to resolve the boundary layers. All simulations were performed on a grid of $N_R \times N_\phi \times N_z \equiv 24 \times 64 \times 24$ nodes.

3 An algorithm for breeding in the rotating annulus

The main use of the breeding method is in operational weather forecasts in a number of countries, including the United States (NCEP, FNMOC) (Tracton and Kalnay, 1993; Toth and Kalnay, 1997; Toth et al., 1997; Pu et al., 1997), Japan (JMA), Korea (KMA), China (CMA), and India (NCMRWF). The method was originally implemented at NCEP using a 17-member ensemble with three control forecasts and seven bred-pair perturbations (Toth et al., 1997, Figs. 1 and 2), and was used there until the Ensemble Transform technique (an extension of the breeding method) was implemented in 2006 (Wei et al., 2008).

Breeding has also been used in other contexts: in El Niño-Southern Oscillation studies (Cai et al., 2003; Yang et al., 2006), in idealized systems like the Lorenz (1963) equations

Table 1. Properties of the annulus and fluid setup. It is identical to the "main comparison" of Hignett et al. (1985).

Annulus inner cylinder radius	a=2.5 cm
Annulus outer cylinder radius	b=8.0 cm
Annulus height	d=14.0 cm
Working fluid (by volume)	83% water / 17% glycerol
Mean density	ρ_0=1.044 g cm^{-3}
Volume expansion coefficient	α=2.86 × 10^{-4}/K^{-1}
Prandtl number	Pr = 13.1
Initial temperature	T_0=20 °C (midway between T_a and T_b)
Viscosity / 10^{-2} cm^2 s^{-1}	ν=1.620$\left[1 - 2.790 \times 10^{-2}(T - T_R) + 6.730 \times 10^{-4}(T - T_R)^2\right]$
Density / g cm^{-3}	ρ=1.043$\left[1 - 3.070 \times 10^{-4}(T - T_R) - 7.830 \times 10^{-6}(T - T_R)^2\right]$
Thermal diffusivity / 10^{-3} cm^2 s^{-1}	κ=1.290$\left[1 + 2.330 \times 10^{-3}(T - T_R)\right]$
	T_R=22° C in each case.

(Evans et al., 2004), in coupled chaotic systems (Kalnay et al., 2003; Peña and Kalnay, 2004; Primo et al., 2005), using a quasi-geostrophic channel model (Corazza et al., 2003), and in a Martian general-circulation model (Newman et al., 2004). Breeding has been used in the context of the rotating annulus by Gilmour (1998, pp.131–136), Smith and Gilmour (1999), and Gilmour et al. (2001), who used a radial basis function model to compare breeding methods and to measure nonlinearity.

The algorithm is adapted from Houtekamer and Derome (1994), and is shown schematically in Fig. 2. In the perfect model scenario, one simulation x^t is defined as the "truth" (unknowable experimentally, due to incomplete and noisy observations), and all the forecasts are then compared with that state.

We shall use $x \equiv x(R, \phi, z, t) \equiv x(\boldsymbol{r}, t) \equiv x(t)$ to denote the complete state of the system as a function of both time and space. x refers to the set of fields u, v, w, T, and Π, all of which are functions of R, ϕ, z, and t. When one of these fields is being considered on its own but in a context applicable to any of the five fields, then $\tilde{x}$ will be used to denote this field.

3.1 Stage zero: initialize simulation

The simulations are initialized by first running a reduced version of MORALS to integrate the axisymmetric form of the equations of motion over a vertical 2D slice (using the same parameters as the subsequent 3D simulation). This is run to a steady state (at around t=10 000 s). The slice is then copied to each azimuthal coordinate, and a sinusoidal perturbation is applied to the temperature field:

$$\delta T(R, \phi, z) \, / \, \text{K} = \begin{cases} 0.1 \sin\left(\frac{R-a}{b-a}\pi\right) \sin\left(\frac{z}{d}\pi\right) & \phi=0 \\ 0 & \text{otherwise} \end{cases}$$

(1)

where a, b, and d are constants defined in Table 1. This perturbation is applied only at ϕ=0, so does not excite any particular azimuthal mode. Each simulation therefore starts from its own axisymmetric state, but is initialized using the same perturbation. The perturbation is necessary to create an azimuthal asymmetry in the model state, as it would otherwise remain axisymmetric (at least until the effects of using finite precision arithmetic grow to appreciable levels).

MORALS is then run from t=0 to t=t_0, to generate the truth state $x^t(t_0)$. t_0 is chosen to be large enough so that any transient oscillations caused by the perturbation in Eq. 1 have decayed.

3.2 Stage one: initialize bred vectors

In stage one, initial perturbations are generated for the forecast stage two ($t_1 \rightarrow t_2$). First, the truth state $x^t(t_0)$ is integrated from $x^t(t_0) \rightarrow x^t(t_1)$.

M+1 simulations x^m, m=[0, ..., M] are then run to generate perturbed states around x^t. One simulation (m=0) will subsequently be used to define the control forecast in stage two, and M simulations (m=[1, ..., M]) will be used to define the perturbed forecasts. In stage one all these simulations are treated identically. Each one is initialized at t=t_0 by adding, at each model grid point, normally distributed random numbers with mean zero and standard deviation $\sigma_{\tilde{x}^t}$ to the truth state $\tilde{x}^t$. This is done in each of the five fields $\tilde{x}^t$=u^t, v^t, w^t, T^t, and Π^t:

$$\tilde{x}^m(t_0) = \tilde{x}^t(t_0) + N(0, \sigma_{\tilde{x}^t}^2) \quad m=[0, ..., M]$$

(2)

Each $\sigma_{\tilde{x}^t}$ is defined as the peak-to-peak value of the azimuthal variation in the field $\tilde{x}^t(t_0)$ at mid-height [z_{mid}=$d/2$] and mid-radius [R_{mid}=$(a+b)/2$], multiplied by some constant

factor F (we use $F=0.1$ in this work – see Sect. 3.4).

$$\sigma_{\tilde{x}^{\mathrm{t}}}(t_0) = F\left[\max_j \tilde{x}^{\mathrm{t}}\left(R_{\mathrm{mid}}, j, z_{\mathrm{mid}}, t_0\right) \right.$$
$$\left. - \min_j \tilde{x}^{\mathrm{t}}\left(R_{\mathrm{mid}}, j, z_{\mathrm{mid}}, t_0\right)\right] \tag{3}$$

These $M + 1$ complete model states are then integrated forward using MORALS for a time Δt_0 (the "rescaling time"). At $t=t_0 + \Delta t_0$, the difference field is calculated between each of the $M + 1$ perturbed states and the truth:

$$x^{\mathrm{d},m}(t_0 + \Delta t_0) = x^m(t_0 + \Delta t_0) - x^{\mathrm{t}}(t_0 + \Delta t_0)$$
$$m=[0, ..., M] \tag{4}$$

These difference fields are then renormalized. The norm is a volume-weighted pressure difference $\eta^m(t)$. In this work, area and volume-weighted statistics will be used a number of times; the weighting is necessary because the spacing between model grid points is non-uniform. For the pressure difference field $\Pi^{\mathrm{d},m}$,

$$\eta^m(t) = \sqrt{\frac{\sum_{ijk}\left(\Pi^{\mathrm{d},m}_{ijkt}\right)^2 \Delta V_{ijk}}{\sum_{ijk}\Delta V_{ijk}}} \tag{5}$$

where ΔV_{ijk} is the volume element at (i, j, k). For each m in turn, $\eta^m(t_0+\Delta t_0)$ is calculated and compared with the equivalent η^m calculated at $t=t_0$. If the perturbation has grown, then

$$\frac{\eta^m(t_0)}{\eta^m(t_0 + \Delta t_0)} < 1 \tag{6}$$

If this is the case, then all five difference fields for ensemble member m are scaled by this ratio to produce a set of **bred vectors**:

$$x^{\mathrm{bv},m}(t_0 + \Delta t_0) = \frac{\eta^m(t_0)}{\eta^m(t_0+\Delta t_0)} x^{\mathrm{d},m}(t_0 + \Delta t_0) \tag{7}$$

If the perturbation has decayed, then

$$\frac{\eta^m(t_0)}{\eta^m(t_0 + \Delta t_0)} \geq 1 \tag{8}$$

In this case, the fields are not rescaled, and the bred vector for ensemble member m is simply the difference field [this renormalization condition is also used by Newman et al. (2004)]:

$$x^{\mathrm{bv},m}(t_0 + \Delta t_0) = x^{\mathrm{d},m}(t_0 + \Delta t_0) \tag{9}$$

This point marks the end of the first breeding cycle. The $M + 1$ bred vectors are then used as new perturbations to the truth state x^{t} at $t=t_0 + \Delta t_0$, and each state is integrated again using MORALS. The cycle is repeated in intervals of Δt_0, comparing the new difference fields with $\eta^m(t_0)$ each time, until the required number of breeding cycles are completed, at $t=t_1=t_0 + N_1\Delta t_0$. (For the choice of N_1, see Sect. 4.)

3.3 Stage two: forecast

Stages zero and one are essentially preparation for stage two, in which the true state x^{t} is forecast using the ensemble perturbed by breeding vectors. We have already defined the truth state $x^{\mathrm{t}}(t_1)$ and $M+1$ bred vectors $x^{\mathrm{bv},m}(t_1)$, $m=[0, ..., M]$, at $t=t_1$. The first step in stage two is to rescale the bred vectors to make their $\eta^m(t_1)$ values equal to F times the peak-to-peak pressure variation in x^{t} over the mid-height/mid-radius azimuthal circle at $t=t_1$:

$$x^{\mathrm{bv},m'}(t_1) = \frac{\sigma_{\Pi^{\mathrm{t}}}(t_1)}{\eta^m(t_1)} x^{\mathrm{bv},m}(t_1) \quad m=[0, ..., M] \tag{10}$$

where $\sigma_{\Pi^{\mathrm{t}}}(t_1)$ is calculated using Eq. (3) and $\eta^m(t_1)$ using Eq. (5). Using $F=0.1$, this rescaling means that the magnitude of perturbations is 10% of the size of the attractor (see Sect. 3.4).

The first bred vector $x^{\mathrm{bv},0'}(t_1)$ is used to define the analysis state, which is the initial condition for the control forecast:

$$x^{\mathrm{c}}(t_1) = x^{\mathrm{t}}(t_1) + \lambda\, x^{\mathrm{bv},0'}(t_1) \tag{11}$$

where λ is the "analysis error amplitude", a constant to be defined shortly. This step is justified by Houtekamer and Derome (1994) and Cai et al. (2003); they argue that in the perfect model scenario bred vectors reflect the difference between truth and analysis.

The other M bred vectors $x^{\mathrm{bv},m'}(t_1)$ with $m=[1, ..., M]$ define an ensemble of $2M$ bred-pair perturbations $x^{\mathrm{B},m\pm}(t_1)$ around the analysis $x^{\mathrm{c}}(t_1)$:

$$x^{\mathrm{B},m+}(t_1) = x^{\mathrm{c}}(t_1) + \mu\, x^{\mathrm{bv},m'}(t_1) \tag{12}$$
$$= x^{\mathrm{t}}(t_1) + \lambda\, x^{\mathrm{bv},0'}(t_1) + \mu\, x^{\mathrm{bv},m'}(t_1)$$

$$x^{\mathrm{B},m-}(t_1) = x^{\mathrm{c}}(t_1) - \mu\, x^{\mathrm{bv},m'}(t_1) \tag{13}$$
$$= x^{\mathrm{t}}(t_1) + \lambda\, x^{\mathrm{bv},0'}(t_1) - \mu\, x^{\mathrm{bv},m'}(t_1)$$

where μ is the "bred vector amplitude", another constant to be defined shortly. This step ensures that the ensemble is centred around the analysis state, and is justified because the distribution of analysis error is assumed to be symmetrical, so addition and subtraction are equivalent.

The $2M+2$ states thus defined at $t=t_1$ are integrated using MORALS from $t_1 \rightarrow t_2 = t_1 + t_{\mathrm{f}}$, which is the end of the forecast. At $t=t_2$, the truth $x^{\mathrm{t}}(t_2)$ is forecast using the control forecast $x^{\mathrm{c}}(t_2)$ and the $2M$ perturbed forecasts $x^{\mathrm{B},m\pm}(t_2)$, for $m=[1, ..., M]$.

3.4 Parameter selection

Two parameters determine the behaviour of a breeding cycle: the rescaling time and the initial amplitude (Kalnay, 2003, Fig. 6.5.3). A third parameter, the initial random seed, distinguishes between individual bred vectors, but in this section we are concerned with parameters that affect the ensemble as a whole.
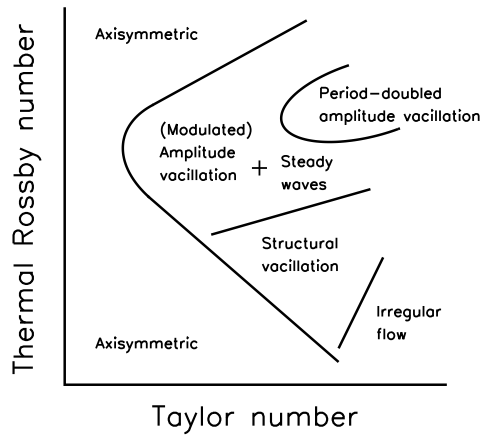
Fig. 3. Regime diagram showing the approximate distribution of annulus flow regimes as a function of the two standard dimensionless parameters, the Taylor number and the thermal Rossby number. This distribution of regimes is seen for the fluid and annulus dimensions used here, in Hignett et al. (1985), and in Young and Read (2008).

The rescaling time (Δt_0) should be short enough so that nonlinear saturation does not occur within one cycle, as otherwise information on the shape and magnitude of the perturbation is lost. Nonlinear saturation occurs if perturbations grow to the size of the system during one breeding cycle, or

$$g_c^m \geq 1/F \tag{14}$$

where F is defined in Sect. 3.2, and calculated below. g_c^m is the "growth factor" – a measure of bred vector growth during cycle c:

$$g_c^m = \frac{\eta^m(t_0 + c\Delta t_0)}{\eta^m(t_0)} \tag{15}$$

The rescaling time should also be long enough for perturbations to become nonlinear (in order to cover the space of possible perturbations as much as possible). Operational centres use several hours (Toth and Kalnay, 1993), but the equivalent in the annulus is $\pi/2\Omega$ (1/4 revolution $\equiv$6 h); this is always too short for interesting behaviour to appear. Gilmour et al. (2001) define the relative nonlinearity and the anticorrelation of a pair of bred vectors; conditions on both must be satisfied for a pair to be linear.

Tests with a number of different parameter combinations showed that Δt_0=70 s is a suitable rescaling time, as it satisfies the condition for nonlinearity, but does not saturate.

The initial amplitude (F) represents an estimate of the size of actual analysis errors. Houtekamer and Derome (1994) test different amplitudes and show that the forecasting advantage in using a breeding ensemble over a control forecast is maximized when this amplitude is about 10% of the natural variability.

Tests were done with varying F. We want the bred vector growth to be as large as possible, so as to sample the largest amount of state space, but without saturating the perturbations in one cycle. Values of F between 0.003 and 0.3 were tested. For F=0.3 some error saturation was observed, so F=0.1 was chosen for use in practice. Co-incidentally, this is the same as Houtekamer and Derome (1994) above.

The analysis error amplitude λ and the bred vector amplitude μ represent the error in the analysis state $x^c(t_1)$, and the initial spread of the ensemble, respectively. They were chosen to maximize bred vector growth over time, and to offer the best improvement in forecast accuracy over the control forecast. Tests using a range of different values gave λ=1.0 and μ=0.5 as suitable values. In each of these tests, the initial growth rate of the bred vector was the same over the whole range of λ and μ used (0.25$\leq\lambda\leq$2.00 and 0.50$\leq\mu\leq$1.00); only after 4–6 cycles did the growth rate begin to diverge[2]. This indicates that the same modes were excited over the whole range of perturbation magnitudes. In the atmosphere, convective modes are preferentially excited at small perturbation magnitudes (Kalnay, 2003, Fig. 6.5.6), so our tests indicate that such convective modes may not be as important in the annulus as they are in the atmosphere. Internal gravity waves are important small-scale modes in the annulus, but from our tests we see no evidence that these are excited over the range of magnitudes used. Therefore, we feel that our conclusions should be robust over a range of initial perturbation magnitudes.

4 Forecast results and analysis

The annulus displays a wide range of behaviour depending on the rotation rate, temperature forcing, fluid, and tank dimensions (Fig. 3). We shall not describe these different dynamical regimes in detail here; see Hide and Mason (1975) or Read et al. (1992) for reviews. Ten complete forecasts were made using the method in Sect. 3, and the forecast parameters are detailed in Table 2. The full range of regimes detailed in Hignett et al. (1985) and Young and Read (2008) were covered, except steady wave flow and irregular flow.

For our purposes, the regimes can be divided into two distinct classes: quasi-periodic and chaotic. From the quasi-periodic regimes, we shall examine one case from the amplitude vacillation (AV) regime, and two from the modulated amplitude vacillation (MAV) regime, both with dominant wavenumber three. We shall examine two chaotic regimes: the wavenumber three structural vacillation regime (3SV), which is believed to display high-dimensional chaotic behaviour (Read et al., 1992), and a wavenumber two "period-doubled" regime (2AV–dh) whose amplitude modulation bi-

---

[2]Young, R. M. B.: Breeding vectors in the rotating annulus as a measure of intrinsic predictability, unpublished internal report, Atmospheric, Oceanic and Planetary Physics, University of Oxford, 68 pp., 2006.

Table 2. Parameters defining each of the forecasts.

| Forecast	A	B	C	D	E	F	G	H	I	J
Regime	AX	3AV	3MAV	3MAV	2AV–dh	2AV–dh	2AV–dh	3SV	3SV	3SV
Parameters defining the flow:										
$\Omega/\,\mathrm{rad\,s}^{-1}$	0.19	0.65	0.81	0.75	0.84	0.845	0.8251	2.50	3.00	3.50
$\Delta T/°\mathrm{C}$	0.01	3.00	3.50	3.00	9.54	9.64	9.21	4.00	4.00	4.00
$\Delta t/\mathrm{s}$	0.05	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01
t_0/s	8000	1500	1500	2000	1500	1500	1000	1500	1500	1500
t_1/s	8700	2200	2200	2700	2200	2200	1700	2200	2200	2200
t_f/s	–	1050	3150	1050	1050	1050	1050	1050	420	2100
$t_1 + t_\mathrm{f}/\mathrm{s}$	–	3250	5350	3750	3250	3250	2750	3250	2620	4300
Derived quantities:										
Taylor number / 10^6	0.188	2.21	3.42	2.94	3.68	3.73	3.55	32.6	47.0	63.9
Thermal Rossby number	0.036	0.92	0.69	0.69	1.76	1.76	1.76	0.083	0.058	0.042
Rotation period / s $=2\pi/\Omega$	33.1	9.67	7.76	8.38	7.48	7.44	7.62	2.51	2.09	1.80
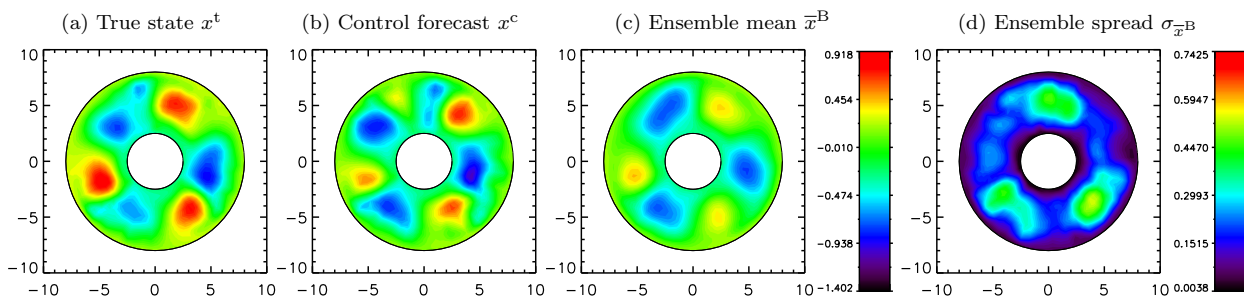
Fig. 4. Example pressure fields from forecast I (3SV) showing different components of the forecast at mid-height (z=7cm) at the end of the forecast (t=420 s). The axes indicate the distance in cm from the rotational axis. The colour scale is the same for Fig. 4a–c, and all scales are in cm^2 s^{-2}.

Table 3. Parameters common to each forecast.

Analysis error amplitude	λ	1.0
Bred vector amplitude	μ	0.5
Breeding cycles in stage one	N_1	10
Stage one perturbation scaling factor	F	0.1
Stage one rescaling time/s	Δt_0	70
Total perturbed ensemble members	$2M$	10

furcates like the logistic map when the Taylor number is varied (Young and Read, 2008). Some of our analysis will focus on three forecasts in particular, one from each of the three main regimes: forecast C (3MAV), forecast G (2AV–dh), and forecast I (3SV).

As a null test of the procedure, one forecast (forecast A) was done in the axisymmetric regime; all forecasts remained within machine precision of the truth. This forecast is therefore omitted from the analysis below, but is included in Table 2 for completeness.

Table 3 lists the parameters common to each forecast. Each forecast used M=5 for a perturbed ensemble of ten members in the forecast stage. This figure is based on the available computer power and the conclusions of Toth and Kalnay (1997, p.3313), who show that much of the predictive advantage gained from an ensemble is obtained with ensembles of size ten. Δt is the model timestep and is chosen to be as large as possible without making the discretized model equations unstable. Ten breeding cycles in stage one (N_1=10) was sufficiently long for coherent structures to emerge in the breeding vectors. The forecast time t_f was chosen to cover several vacillation cycles.

Typical pressure fields from the forecasts are shown in Fig. 4; this figure shows pressure fields at mid-height at the end of the 3SV forecast with Ω=3.00 rad s^{-1} and ΔT=4.0°C (forecast I). During stage two, the ensemble mean is the average of the $2M$ perturbed forecasts:

$$\bar{x}^{\mathrm{B}}(t) = \frac{1}{2M} \sum_{\substack{m=M- \\ m \neq 0}}^{M+} x^{\mathrm{B},m}(t) \tag{16}$$
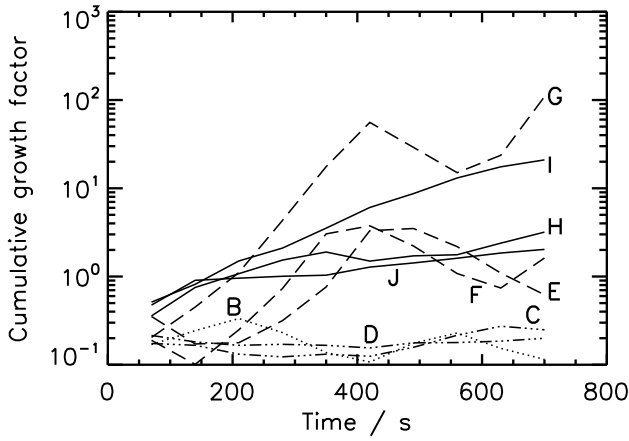
Fig. 5. Bred vector cumulative growth during the breeding vector initialization stage. Each line shows the mean cumulative growth (averaged over the bred vectors) as a function of time from the start of this stage. Each breeding cycle is 70 s long. The 3AV forecast is indicated by a dotted line, 3MAV by dot-dashed lines, 2AV-dh by dashed lines, and 3SV by solid lines. Each letter corresponds to a forecast in Table 2.

and the forecast error variance is

$$\sigma_{\bar{x}^{\mathrm{B}}}^2(t) = \frac{1}{2M} \sum_{\substack{m=M- \\ m \neq 0}}^{M+} \left[x^{\mathrm{B},m}(t) - \bar{x}^{\mathrm{B}}(t) \right]^2 \qquad (17)$$

For ease of comparison, most of the analysis below uses the pressure field at mid-height.

4.1 Bred vector growth rates

We shall first examine two quantitative measures of predictability made during stage one. These give us an a priori estimate of the predictability in the second stage (at least in principle). In this and Sect. 4.2 all the times referred to are relative to the start of stage one.

We define the cumulative growth factor for breeding vector m after cycle C as

$$G_C^m = \begin{cases} g_1^m & C = 1 \\ g_C^m G_{C-1}^m & g_{C-1}^m \geq 1 \\ \frac{g_C^m}{g_{C-1}^m} G_{C-1}^m & g_{C-1}^m < 1 \end{cases} \qquad (18)$$

Less formally, G_C^m is g_C^m multiplied by the cumulative growth the last time the fields were rescaled. This formulation is required because the field is only rescaled if the breeding vector grows during the cycle.

Figure 5 shows, for each forecast, the mean cumulative growth factor $\overline{G_C}$ as a function of time in stage one, averaged (geometric mean) over the $M+1$ bred vectors. The error bars are omitted for clarity; they are in the range 0.2–0.8.

The 3AV forecast is indicated by a dotted line, 3MAV by dot-dashed lines, 2AV–dh by dashed lines, and 3SV by solid lines. The letters correspond to the forecasts in Table 2.

As the plot is logarithmic, straight lines with positive gradient indicate exponential growth of the breeding vector. This is seen in the 3SV forecasts (solid lines). The approximately flat lines for 3AV and 3MAV indicate that these perturbations decay by a factor of five during the first cycle, and then remain at that size over the remainder of stage one (as the perturbations are not rescaled if $g_C^m < 1$, from Eqs. 8–9).

The behaviour of the 2AV–dh forecasts are quite different, however. There are large fluctuations in the cumulative growth plots themselves, and a large range in the final cumulative growth values between the three forecasts, compared with the other regimes. During the first half of the stage, the cumulative growth increases but during the second half it decays again, before starting to grow during the final cycles. By comparing the cumulative growth with a time series showing the magnitude of the dominant pressure wavenumber over the same time frame, it was found that this oscillation is strongly correlated with the vacillation cycle. Both have approximately the same period ($\sim$400 s), and the perturbation growth rate is highest as the amplitude of the dominant wave increases. This same behaviour is seen in all three 2AV–dh forecasts, and also 3AV. It is not seen in either 3MAV or 3SV, possibly because the amplitude of the dominant wave does not change appreciably over time.

4.2 Bred vector dimension

The bred vector dimension Ψ is a measure of the local dimension of a set of breeding vectors (Patil et al., 2001). Francisco and Muruganandam (2003) show that for a system of coupled Lorenz equations, Ψ is directly related to the intrinsic predictability of a system: the higher the value of Ψ, the less predictable the system. We shall test whether this is also the case in the perfect model rotating annulus (a much more complex system), by calculating a measure of Ψ for each of the forecasts during stage one, and comparing with later measures of predictability. The bred vector dimension is defined at each grid point as follows:

$$\Psi(\sigma_1, \sigma_2, ..., \sigma_L) = \frac{\left(\sum_{l=1}^L \sigma_l \right)^2}{\sum_{l=1}^L \sigma_l^2} \qquad (19)$$

We follow Patil et al. (2001) to calculate the σ_l. The σ_l are the singular values of a $50 \times L$ matrix constructed from L column vectors of length 50. The elements of the l–th column vector are the u and v velocities of the l–th bred vector at the 25 nearest horizontal grid points (arranged in a 5×5 box). A box of 5×5 grid points is appropriate here, as it corresponds to a box size of about $1\,\mathrm{cm} \times 1\,\mathrm{cm}$ near the inner and outer walls, and $2\,\mathrm{cm} \times 2\,\mathrm{cm}$ in the interior of the fluid; this is about the size of medium-large scale features in the flow. The u and v data are first defined as two separate vectors of

length 25, normalized, and the u velocity is rescaled to give both u and v the same mean square norm. The two vectors are combined into one vector of length 50, which is again normalized to length one. This process is repeated for all the ensemble members $x^{\mathrm{bv},m}$, $m=[0, ..., M]$, then combined to form a $50 \times L \equiv 50 \times (M+1)$ matrix, and Ψ is then calculated using Eq. (19). Ψ can take values between 1 and $L = M+1$ ($=6$ here), and we use the two components of horizontal velocity because they were used in previous work with the bred vector dimension (Patil et al., 2001).

For each of the forecasts, Ψ was calculated at each grid point after each cycle during stage one. For a simple comparison between forecasts, the mean $\overline{\Psi}$ was calculated for the entire fluid:

$$\overline{\Psi}(t) = \frac{1}{V} \int_V \Psi(\boldsymbol{r}, t) dV \qquad (20)$$

Figure 6 shows this quantity over time for each forecast (NB: To produce this diagram, stage one was repeated for each forecast, for $N_1 = 20$ instead of $N_1 = 10$). These results show clear differences between the three regimes. Initially $\overline{\Psi}$ begins at high values, as the initial perturbations are random. Over time, as the breeding vectors become more structured, we expect the value of $\overline{\Psi}$ to equilibrate. For 3SV (solid lines in Fig. 6), $\overline{\Psi}$ does not equilibrate, but varies between 3.2 and 4.2. For 2AV–dh, all three cases (dashed lines) eventually fall to between 1 and 2, which is a surprising result, as this is a chaotic regime. The quasi-periodic cases (dotted and dotdash lines) do not fall to one, but equilibrate in the range 2.2–3.2. From Francisco and Muruganandam (2003), we would expect $\overline{\Psi}$ for the chaotic regimes to be higher.

These results predict that 3SV will be the least predictable regime during the forecast stage, and 2AV–dh will be the most predictable (a surprising result).

We shall use the results from both Sects. 4.1 and 4.2 later, as predictors of predictability.

4.3 Error variance

We now turn to analysis of the forecast stage. In the rest of Sect. 4, references to time are relative to the start of the forecast stage ($t=t_1$).

One of the most commonly used quantitative measures of the progress of a forecast is the error variance. The mean error variance between a forecast and the truth as a function of time for the pressure field Π is

$$s(t)^2 = \frac{\sum_{ijk} \left(\epsilon_{ijk} - \bar{\epsilon} \right)^2 \Delta V_{ijk}}{\sum_{ijk} \Delta V_{ijk}} \qquad (21)$$

where

$$\bar{\epsilon}(t) = \frac{\sum_{ijk} \epsilon_{ijk} \Delta V_{ijk}}{\sum_{ijk} \Delta V_{ijk}} \qquad (22)$$

and

$$\epsilon_{ijk} = \Pi(i, j, k, t) - \Pi^{\mathrm{t}}(i, j, k, t) \qquad (23)$$

ϵ_{ijk} is the difference between the two fields at spatial coordinates ijk, and ΔV_{ijk} is the volume of the grid element over which it is valid. Higher s^2 indicates a poorer forecast. For each forecast, s^2 is plotted as a function of time in Fig. 7 for both the ensemble mean $\bar{x}^{\mathrm{B}}(t)$ (solid line, Eq. 16) and the control forecast $x^{\mathrm{c}}(t)$ (dashed line). Two conclusions can be drawn. First, in almost all cases the ensemble mean performs better than the control. Second, there is a clear difference between the quasi-periodic regimes (Fig. 7a–c), and the chaotic regimes (Fig. 7d–i): The error variance for the AV and MAV forecasts is very small, and not increasing (forecasts B and C are slowly decreasing, in fact). This indicates that the forecasts are very good in these regimes, as one might expect – the model trajectories are falling onto non-chaotic attractors, and hence their behaviour may be indefinitely predictable. The chaotic forecasts are very different: the error variance is some two or three orders of magnitude larger than the quasi-periodic forecasts, in general. The 2AV–dh forecasts have generally increasing trends, and there are large fluctuations in the plots, like in the bred vector cumulative growth (again, on the order of the vacillation period). The 3SV forecasts also have much larger error variance than the quasi-periodic forecasts, but only forecast J is increasing over time. The fact that the error variances in forecasts H and I have stopped increasing may imply that their predictability limit has been reached after a few hundred seconds.

In conclusion, these results give preliminary indications that the 2AV–dh and 3SV regimes are much less predictable than the 3AV and 3MAV regimes.

4.4 Behaviour at a single point

An intuitive way to look at predictability is the behaviour of the ensemble at a single point. Figure 8 shows the pressure at mid-radius/mid-height, at $\phi=0.147$ rad, for each of the nine forecasts. The red lines indicate the perturbed ensemble forecast, the green line is the control forecast, and the black line is the true state.

In the 3AV and 3MAV forecasts (Fig. 8a–c), the ensemble is tightly bound for the whole forecast, again indicating that these regimes may be indefinitely predictable.

The 2AV–dh forecasts (Fig. 8d–f) are well forecast for a time, before the ensemble splits suddenly. The ensemble members in forecast F split into two groups around $t=700$ s, and forecast G first splits around $t=400$ s. The ensemble in forecast E begins to spread around $t=100$ s, and eventually splits into two groups around $t=1000$ s. The next section will examine this behaviour in more detail.
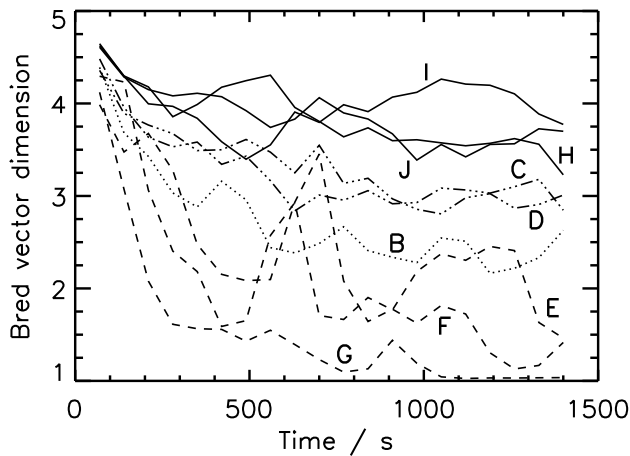
Fig. 6. Bred vector dimensions in extended re-runs of the breeding vector initialization stage. Each line shows the volume-weighted mean bred vector dimension $\overline{\Psi}$ as a function of time, measured from the start of this stage. The line styles and letters are the same as in Fig. 5.

The 3SV forecasts (Fig. 8g–i) show the quickest initial spread. However, after the first few hundred seconds the spread of values remains approximately constant; compare this with the error variances in Fig. 7, which also plateau after a few hundred seconds. The range of pressure values over the whole domain is 4–6 cm^2 s^{-2}, so we might expect the forecast spread to approach this range over time. However, in Fig. 8 the range is about 1 cm^2 s^{-2}. For a chaotic system, the spread should be exponential until the attractor size is reached. Therefore, it is possible that SV is a secondary instability which saturates at a low amplitude relative to the primary (baroclinic) instability; this may lead to a "fuzzy manifold" on the baroclinic wave.

The ensembles in the 3SV regime follow the truth (at this location at least) better than the period-doubled regime. The period-doubled forecasts are affected by a number of "forecast busts", where the truth state lies outside both the groups, even after the ensemble has split. This happens particularly in forecast G, around t=350–550 s, and 700–900 s. By contrast, the truth state remains within the 3SV ensembles at all times except for two short periods during forecast I ($t \sim 230$ s) and forecast J (t=400–500 s).

4.5 Spaghetti plots

The top row of Figs 9–11 show forecasts in the 3SV, 2AV–dh, and 3MAV regimes, using a visualization technique commonly called a "spaghetti plot" in weather forecasting (Tracton and Kalnay, 1993). A single pressure contour is plotted for each perturbed ensemble member (red), as well as the control (green), truth (blue), and ensemble mean (black), at mid-height. The contour is determined by the pressure range

at that level: the contour used is the median of the range over the whole forecast.

There is a qualitative difference in the predictability of the flows. The SV forecast (Fig. 9) spreads out over time, and predictability is lost over a timescale of hundreds of seconds. The period-doubled forecast (Fig. 10) is also unpredictable on a similar timescale, but loses its predictability in a different way. The ensemble behaves similarly to the quasi-periodic forecast (Fig. 11) for the first part of the forecast, but as the amplitude of the wave goes through its minimum (between t=300 s and t=400 s in the figure), the ensemble splits into two groups of approximately equal size. In one group, the peaks of the waves before the minimum remain as peaks after the minimum; in the second, the peaks before the minimum become troughs afterwards. The two groups are π out of phase after the minimum. This behaviour is reminiscent of the Lorenz (1963) attractor, where the unstable fixed point at the origin causes trajectories to diverge unpredictably between the two wings of the attractor.

On further inspection, it was found that there is some correlation between the groups in the period-doubled forecasts and whether the perturbation applied was positive or negative at the start of the forecast stage. In forecast G, the two groups are split by original positive/negative perturbation for the entire length of the forecast (the negative perturbations follow the truth state). In forecast E this is also the case initially, but by the end of the forecast the groups have mixed (the positive perturbations follow the truth in this case, however). This is also the case for forecast F, where one group consists only of two ensemble members and the truth state itself. Finally, the split does not always occur at the first minimum. In forecast E it occurs on the third minimum, and in forecast F on the second (this can also be seen in Fig. 8d–e).

3MAV flow (Fig. 11) is very predictable, as all the ensemble members and the truth are bunched together for the entire length of the forecast (for this reason, only the final timestep is shown here). This behaviour also occurred for 3AV flow (forecast B, not shown).

4.6 Probability forecasts and Brier score

We can quantify the spaghetti plots presented above by considering the ensemble as a probability forecast. The bottom row of Figs 9–11 show probability forecasts for the three forecasts presented above as spaghetti plots. Each plot shows, at each point at mid-height, the percentage of perturbed ensemble members whose pressure values at that level are above the contour values used to define the spaghetti plots in those figures. Suppose, by analogy with weather forecasting, that a particular event is conditional on the pressure being above this value. The plots on the bottom row of Figs 9–11 give the forecast probability that the event will occur. Again, the 3MAV case (Fig. 11) is the most predictable: whether the event will or will not occur is predicted with near-certainty at almost all points. The 2AV–dh and
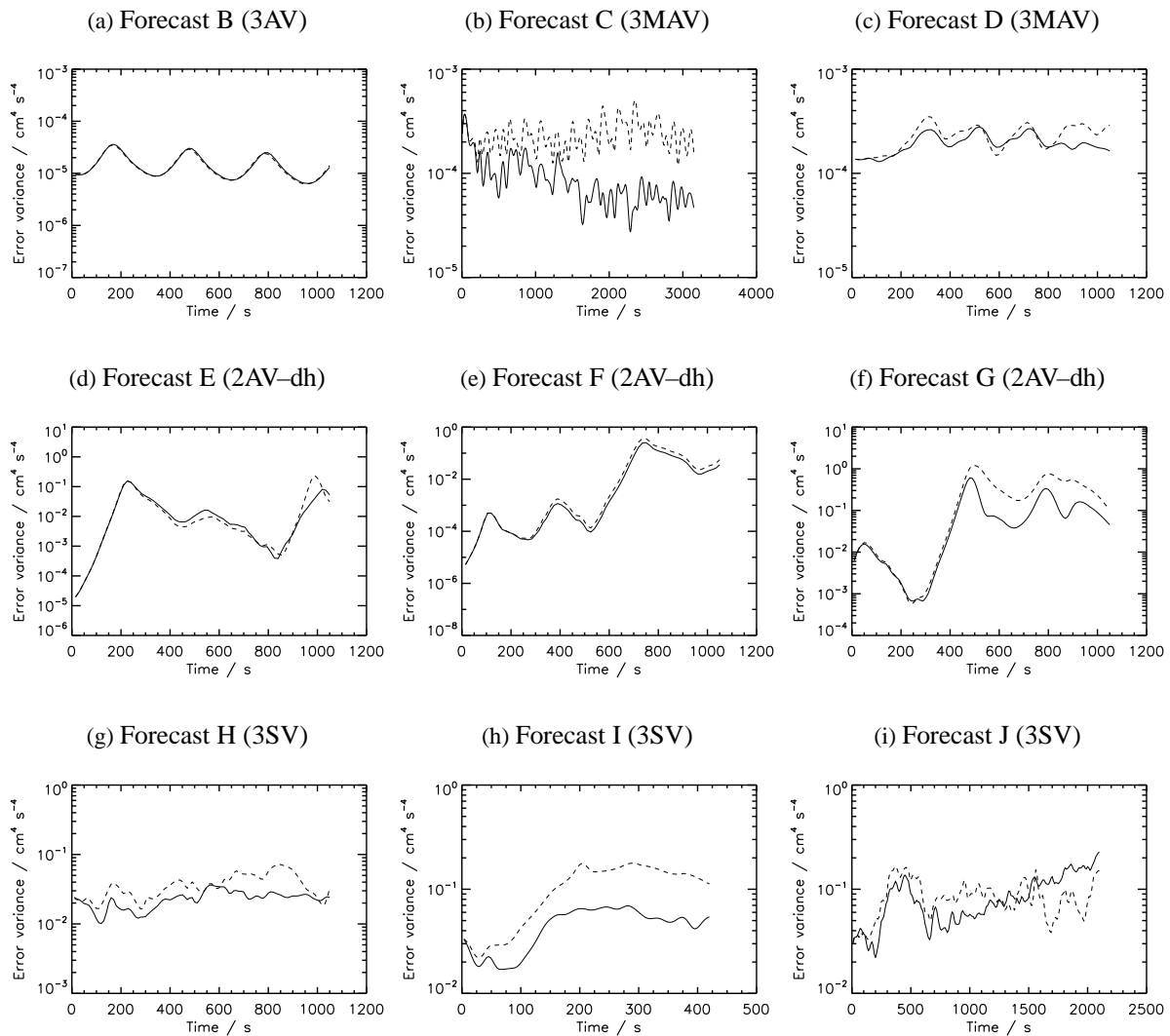
(a) Forecast B (3AV)

(b) Forecast C (3MAV)

(c) Forecast D (3MAV)

(d) Forecast E (2AV–dh)

(e) Forecast F (2AV–dh)

(f) Forecast G (2AV–dh)

(g) Forecast H (3SV)

(h) Forecast I (3SV)

(i) Forecast J (3SV)

Fig. 7. Mean error variance over the whole pressure field (Eq. 21) as a function of time during each forecast. The solid line shows the volume-weighted mean error variance for the ensemble mean, and the dashed line shows the equivalent for the control forecast.

3SV forecasts, however, show a significant drop in predictive power over the course of the forecast. The ensemble splitting in the 2AV–dh forecast (Fig. 10, $t > 400$ s) results in a significant loss of predictive power, as only about 50% probability is assigned to each of the two groups – not much better than guessing one way or the other with no knowledge about the system at all. In the 3SV forecast (Fig. 9), the initial certainty in the forecasts at the wave peaks is quickly lost as the ensemble members spread out over the domain.

These probability forecasts may be evaluated using the Brier score (Brier, 1950). This quantifies the success of the probability forecast in predicting whether the "event" above occurs. At each grid point and time, if the forecast assigns the probability f_1 that the event will occur, and the probability $f_2 = 1 - f_1$ that it will not, the Brier score is

$$B(t) = \sum_{j=1}^{2} \left(f_j - E_j \right)^2 = 2 \left(f_1 - E_1 \right)^2 \qquad (24)$$

where $E_1 = 1$ if the event does happen, and $E_1 = 0$ if it does not. B takes values in the range 0–2, where lower values indicate higher forecast skill.

This quantity was calculated at each grid point in the three forecasts above, and these results are summarized in Fig. 12. The top row shows contours of Brier score at mid-height at the forecast time corresponding to the final frame in the sequence of forecasts in Figs. 9–11. The line plots in the bottom row show the mean Brier score as a function of time, weighted by volume over the whole domain. An increasing Brier score corresponds to decreasing forecast skill.
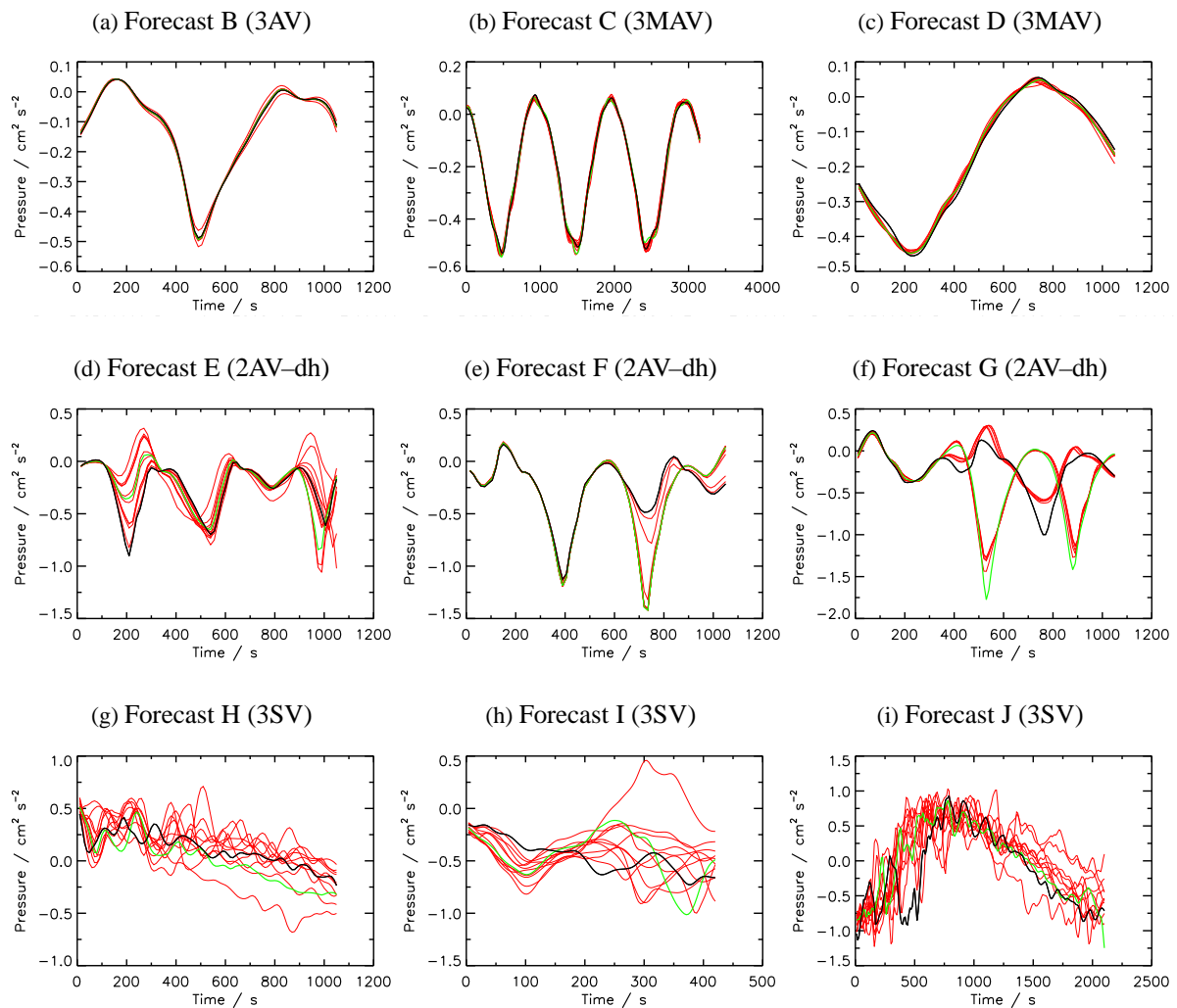
Fig. 8. The behaviour of the ensemble at a single point during the forecast stage. Pressure (Π) is plotted as a function of time at mid-height / mid-radius (R=5.25 cm, z=7 cm), at ϕ=0.147 rad. Red lines show the pressure of the perturbed ensemble members, the green line is the control, and the black line is the truth.

The Brier score for a forecast probability of 50% is 0.5, if there are only two possible outcomes. In general, the line plots in Fig. 12 show that on average, the true value was predicted with better than 50% probability at each point, which is encouraging. It is difficult to say from these plots whether there is a general upwards trend, associated with the forecast getting worse over time. The near-certainty associated with the probabilities in Fig. 11 (forecast C, 3MAV) is well-founded: the Brier score in Fig. 12c is very small, both the volume-weighted values and over the area shown, except for a few places on the boundary between "definitely yes" and "definitely no". The 2AV–dh forecasts (Fig. 12b, forecast G) are severely diminished by the splitting ensemble, resulting in $B\sim1$ over large areas. The spreading of the forecast ensemble in the 3SV case (Fig. 12a) results in the Brier score

being high in regions where the gradient of the pressure field is highest.

In any discussion of predictability it is helpful to ask how "useful" the ensemble is compared to alternative forecasts. A forecast using the climate is one such alternative. The ensemble is then only "useful" as long as it can predict the true state better than the climate. Using the Brier score, the ensemble and the climate can be compared as predictors of the true state. By plotting a cumulative frequency distribution showing the fraction of the fluid volume below each possible score, the distribution of skill from the ensemble can be compared with what is expected from a forecast using the climate to define the probabilities. This representation filters out isolated abnormally poor or good forecasts.
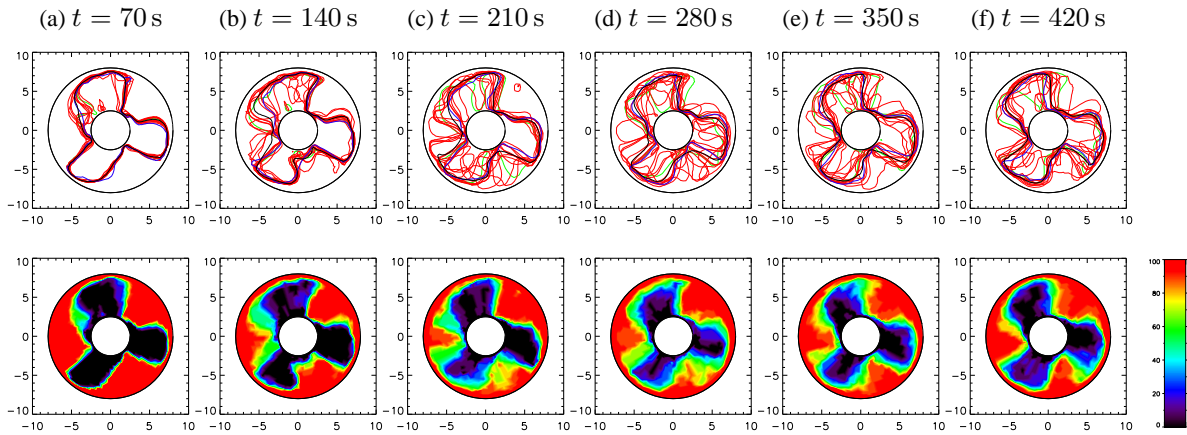
Fig. 9. Top: sequence of spaghetti plots showing an ensemble forecast of pressure in the structural vacillation regime (forecast I). All plots show the $\Pi=-0.201\ \text{cm}^2\text{s}^{-2}$ pressure contour at $z=7$ cm. The axes indicate the distance in cm from the rotational axis. Red contours indicate the perturbed forecasts $x^{B,m\pm}$, green the control forecast x^c, black the ensemble mean $\bar{x}^B$ (Eq. 16) and blue the truth x^t. Bottom: the equivalent probability forecast for the same regime. Each plot shows the percentage of perturbed ensemble members above the pressure contour $\Pi=-0.201\ \text{cm}^2\text{s}^{-2}$, over the horizontal slice at $z=7$ cm. Black is 0% probability, and red is 100%.
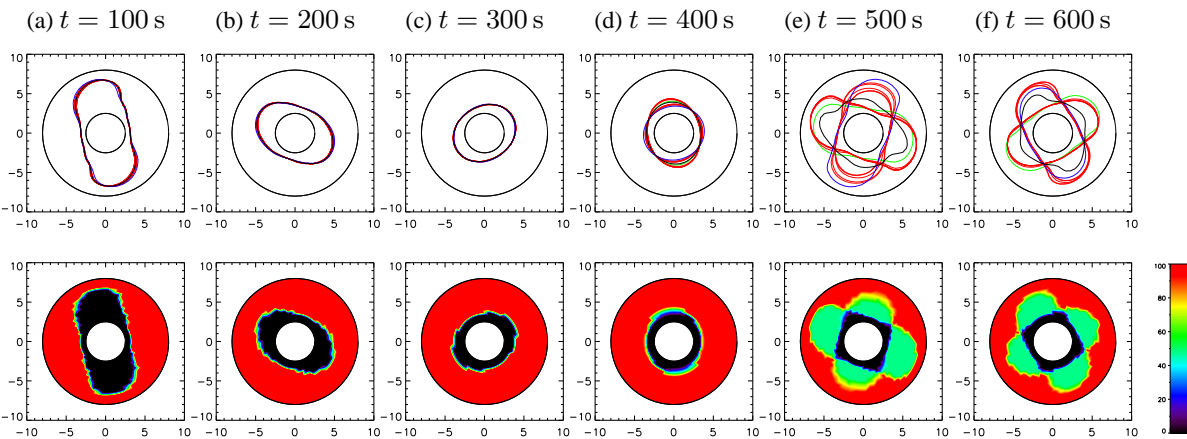
Fig. 10. As Fig. 9, but in the period-doubled regime (forecast G), at $\Pi=-0.561\ \text{cm}^2\text{s}^{-2}$.

To construct a probability forecast using the climate, the climatological pressure distribution is defined at each point in the flow as the distribution of pressure over a long simulation (the truth simulation over stage two is used). The probability that the "event" considered above will occur according to the climatological distribution is then assigned. The Brier score at each point is calculated using the climatological probability as the forecast. At each forecast time, the forecast using the climatology can then be directly compared with the ensemble. We define a cumulative frequency distribution for the fluid volume containing grid points with Brier score b or below, as a function of b:

$$CF(b)=\sum_{B(R,\phi,z)\leq b}\Delta V(R,\phi,z) \qquad (25)$$

This calculation was done for each forecast as a function of time. Figures 13 and 14 show the results at four times during forecasts I (3SV) and G (2AV–dh). The solid lines show the distribution of Brier scores using the climatological distribution as the forecast, and the dashed lines the same for the ensemble forecast. Four times during the forecast are shown, in different colours. If the ensemble is more skilful, the dashed line will be to the left of the solid line, and vice versa.

In general, the climate is a more skilful predictor of the truth for the 3SV forecast in Fig. 13. A useful comparison to make is the value at $B=0.5$. This corresponds to the volume fraction where the true state is predicted with better than 50% probability. In general, the results for forecast I (Fig. 13) are encouraging, as about 80%–90% of points satisfy this
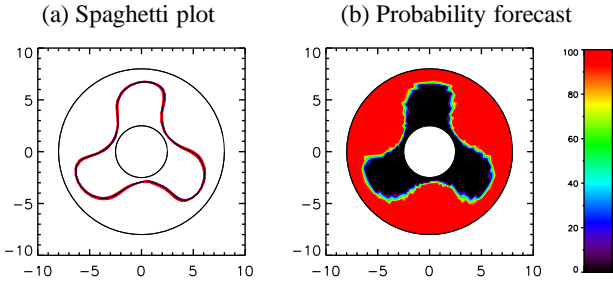
(a) Spaghetti plot   (b) Probability forecast

Fig. 11. As Fig. 9, but in the modulated amplitude vacillation regime (forecast C), for $\Pi = -0.238\,\mathrm{cm}^2\mathrm{s}^{-2}$, at $t = 3150\,\mathrm{s}$ only.

condition, using either the ensemble or the climate as the predictor. Around $t \sim 120\,\mathrm{s}$ there is a transition from the ensemble to the climatology as the better predictor of the true state, and this remains so for the rest of the forecast. This is not the case for the other 3SV forecasts, where the ensemble remains marginally better than the climatology, even at the end.

The 2AV–dh results (Fig. 14) are affected, again, by the ensemble forecast splitting into two. The sharp jump around $b = 0.6$ in all cases but $t = 250\,\mathrm{s}$ in Fig. 14 shows that after 250s about 15–25% of the volume is predicted with 50% accuracy at best.

Figure 15 shows the same as Figs. 13–14 but for forecast C (3MAV), and is representative of the quasi-periodic forecasts: the ensemble forecast is significantly better than the climatology at all times.

4.7 Kolmogorov-Smirnov goodness-of-fit tests

The Kolmogorov-Smirnov (K-S) goodness-of-fit test (Massey, 1951) is another way to estimate the "usefulness" of an ensemble forecast relative to the climate. The usefulness of the forecast is exhausted when the distribution of ensemble forecast values and climatological forecast values can be drawn from the same statistical distribution. At that time, an equally good forecast could be obtained by using the climate as the predictor. Consider the following null hypothesis:

H_0: The cumulative frequency distribution of pressure values predicted by the ensemble and control forecasts at a point in time and space, S, is drawn from the same statistical distribution as the cumulative frequency distribution of the climatological values at that point, F_0.

and the alternative hypothesis:

H_1: S and F_0 are drawn from different statistical distributions.

The K-S test assigns a confidence level for retaining the null hypothesis. The climatological cumulative frequency distribution F_0 is defined as

$$F_0(R, z, x) = \frac{\sum_{n=1}^{N_2} \sum_{j=1}^{N_\phi} \left[\Pi^{\mathrm{t}}(R, \phi_j, z, t_n) \le x\right]}{N_\phi N_2} \tag{26}$$

where N_2 is the number of datasets obtained during the forecast stage, and $N_\phi = 64$ is the number of azimuthal grid points, taking advantage of the rotational symmetry of the system to define a more accurate climatological distribution (note that volume-weighting is not needed here as the azimuthal model grid is uniform). At each point in the vertical slice, F_0 is the fraction of points in the time series at (R, z) below the value x. An equivalent cumulative frequency distribution is calculated at each point in the ensemble forecast, as a function of time:

$$S(R, \phi, z, t, x) =$$

$$\frac{\left[\Pi^{\mathrm{c}}(R,\phi,z,t) \le x\right] + \sum_{\substack{m=M- \\ m \ne 0}}^{M+} \Pi^{\mathrm{B},m}\left[(R, \phi, z, t) \le x\right]}{2M+1} \tag{27}$$

The maximum difference between the two is the K-S statistic D:

$$D(R, \phi, z, t) = \max_x \left| F_0(R, z, x) - S(R, \phi, z, t, x) \right| \tag{28}$$

Each value of D corresponds to a confidence level for retaining the null hypothesis, calculated using Press et al. (1992, pp.617–622).

The K-S statistic D and the corresponding confidence level for retaining H_0 were calculated at each point and time in forecasts B-J. The area-weighted mean $\overline{D}$ was also calculated for each vertical slice. As the ensemble forecast approaches the climatological distribution, $\overline{D}$ will fall to zero. Figure 16 shows the evolution of $\overline{D}$ over the vertical slice at $\phi = 0.147\,\mathrm{rad}$ for each forecast, as a function of time.

The quasi-periodic forecasts (Fig. 16a–c) remain at $\overline{D} > 0.5$ for the whole forecast duration, with no general decreasing trend (except possibly in forecast D). Therefore H_0 is rejected at all points and times with almost 100% significance – the ensemble forecast retains its predictive power at all times.

The two chaotic regimes (Fig. 16d–i) have general downward trends, indicating that $\overline{D}$ should eventually approach zero. Over the forecast period, the value of D where the null hypothesis is retained even at the 50% confidence level is not reached: $D_{50\%} = 0.239$, which is not reached in any of the forecasts. If the rate of decrease remains constant, $\overline{D}$ will fall approximately to zero (and all 'usefulness' in the ensemble forecast will be lost) after about 4000–6000 s for the 2AV–dh forecasts, and after about 1500–3000 s for the 3SV forecasts. These values indicate that the ensemble will remain
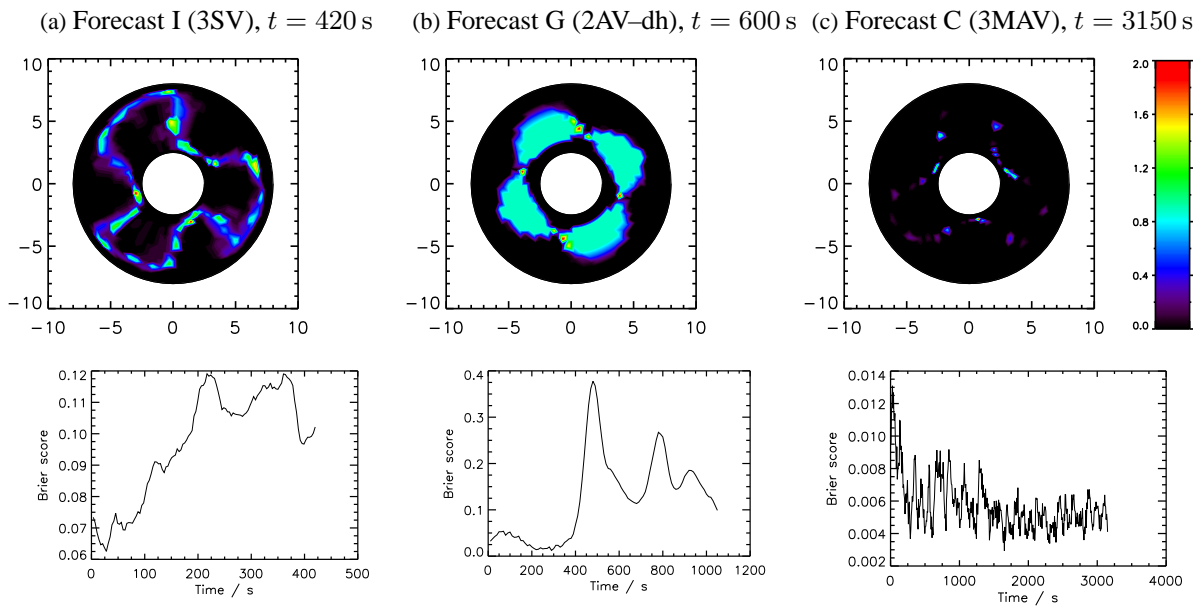
Fig. 12. Top: Horizontal distribution of Brier score in the three forecasts shown earlier as spaghetti plots and probability diagrams. The scores shown in this figure are from the end of the sequences shown in Figs 9–11, at the same height. The colour scale is the same for each diagram, and the axes indicate the distance in cm from the rotational axis. Bottom: volume-weighted mean Brier score as a function of time over the forecast stage, for the same three forecasts.
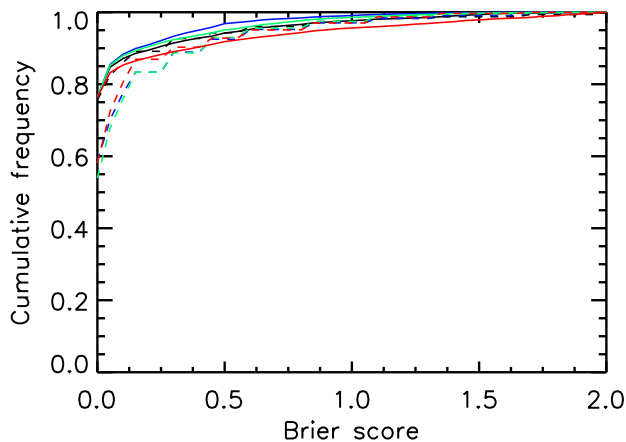

Fig. 13. Brier score cumulative frequency distributions for forecast I (3SV). Each plot shows the fraction of annulus volume where the forecast scores below each Brier score, as a function of Brier score. Four times during the forecast are shown, in different colours. For each colour, the solid line shows the distribution using the climatology as the predictor, and the dashed line shows the distribution using the ensemble. Black $=105$ s, blue $=210$ s, green $=315$ s, and red $=420$ s.

Fig. 14. As Fig. 13, but for forecast G (2AV–dh). Black $=250$ s, blue $=500$ s, green $=750$ s, and red $=1000$ s.

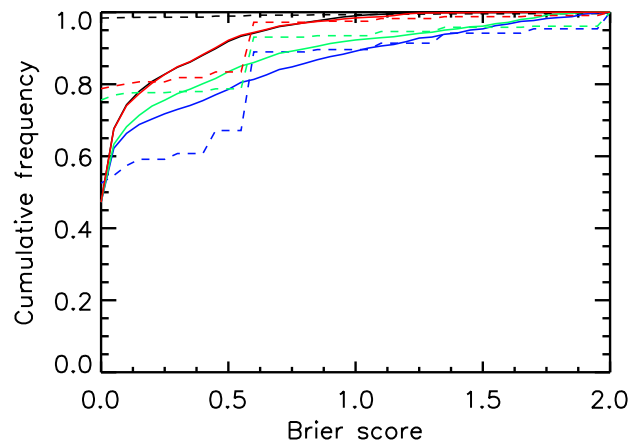more useful for forecasting the truth than the climatological distribution for some time.

5 Summary and conclusions

We have detailed the first part of a study of deterministic predictability in the rotating annulus laboratory experiment. This work is the first study of the annulus to focus on predictability of the first kind (Lorenz, 1975); the second kind of predictability defined by Lorenz is already well characterised by previous studies.
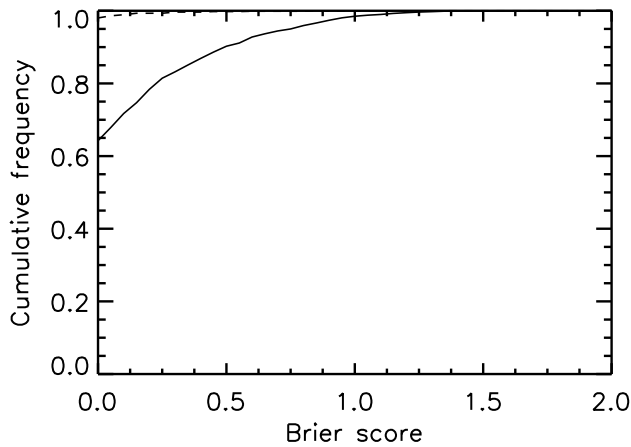
Fig. 15. As Fig. 13, but for forecast C (3MAV), at $t=3150$ s only.

Forecasts of the rotating annulus in the perfect model scenario have shown significant differences between the predictability of quasi-periodic and chaotic flow regimes. Computational evidence suggests that quasi-periodic regimes may be forecast indefinitely using an ensemble prediction scheme, but the predictability of structural vacillation and period-doubling regimes is limited to a timescale of hundreds of seconds. Furthermore, there is a fundamental difference in the way predictability is lost between low-order chaotic regimes (2AV–dh) and high-order chaos (3SV).

Initial estimates of predictability using the bred vector growth rate predicted that quasi-periodic regimes would be much more predictable than chaotic regimes, but it was not possible to make any quantitative separation between the two chaotic regimes. (Note that – strictly – the bred vector growth rate only reflects the dynamics at the selected magnitude of perturbations, so using it to predict the behaviour of the whole system may be limited.) The bred vector dimension gave the surprising prediction that 3SV would be the least predictable and 2AV–dh (which is also chaotic) the most predictable.

Forecasts were made at nine points in parameter space using an ensemble of bred-pair perturbations (Eqs. 12–13), assuming a perfect model. The same technique was applied to each regime, so all things being equal, the differences in the forecasts should reflect the differences between the regimes themselves. Error variance statistics showed a marked forecast improvement by using the ensemble mean instead of a control forecast. Spaghetti plots and probability forecasts showed the chaotic regimes are much less predictable than the quasi-periodic regimes; predictability in the 2AV–dh regime was lost primarily at the minimum point in the vacillation cycle, which was confirmed by examining behaviour at a single point. It would be interesting to see whether this ensemble split also occurs in quasi-periodic flows with a large vacillation index (i.e. where the wave amplitude approaches zero at the cycle minimum).

Comparisons were made between the ensemble forecast and a climatological forecast using the Brier score and the Kolmogorov-Smirnov goodness-of-fit test. The Brier score identified regions of high pressure gradient to be the least predictable parts of the flow, at least in the 3SV regime. We are not aware of any quantitative numerical weather prediction studies that verify this behaviour, so this may be an interesting result worthy of further study. The skill of the 2AV–dh forecasts is diminished by the ensemble split, with some regions of the flow better predicted by the climatology than the ensemble. Kolmogorov-Smirnov tests indicated that forecasts in quasi-periodic regimes may outperform the climatological forecast indefinitely, while the ensemble "usefulness" in the chaotic regimes will be exhausted after a few thousand seconds.

We have used a number of predictability measures in this work, which (apart from the bred vector dimension) all gave the same conclusion about the relative predictability of the regimes studied: 3SV and 2AV–dh are much less predictable than 3MAV and 3AV. The predictions made using bred vector growth rate and dimension hold in all cases except the anomalous bred vector dimension results for the 2AV–dh forecasts. The measures which have enabled us to distinguish most clearly between regimes are behaviour at a point (Fig. 8) and the Brier score – the contour plots in Fig. 12 are a particularly clear indication of which regions of the flow are the most predictable. We noted in the introduction that the predictability of a system should not depend on the method used to measure it. This work has compared a number of validation techniques and predictability measures, and our results indicate that the bred vector dimension may be a questionable measure of predictability in complex systems such as the annulus (or indeed the atmosphere), and that its power as a predictor is limited to low-order models like those used by Francisco and Muruganandam (2003).

Gathering together all the results above, approximate prediction times in the perfect model scenario are summarized below, with equivalent times in annulus rotation periods and Lyapunov times [exponents are taken from Young and Read (2008)]. The values for 3AV/3MAV are lower bounds, as predictability was retained throughout the whole forecast in each case.

Regime	3SV	2AV–dh	3(M)AV
Prediction time / s	300–500	500–1000	>3000
Rotation periods	120–280	65–130	>400
Lyapunov times	1.0–1.5	1.5–3.0	>0.5

Of course, it is not possible to make any firm statements about the predictability of a system without using data from the system itself. We do not claim that the numbers above reflect the true absolute predictability of these regimes – such conclusions cannot be made without using experimental data, in any case, so will need to wait until further work using real data has been completed. In this study, the conclusions about the relative predictability of different
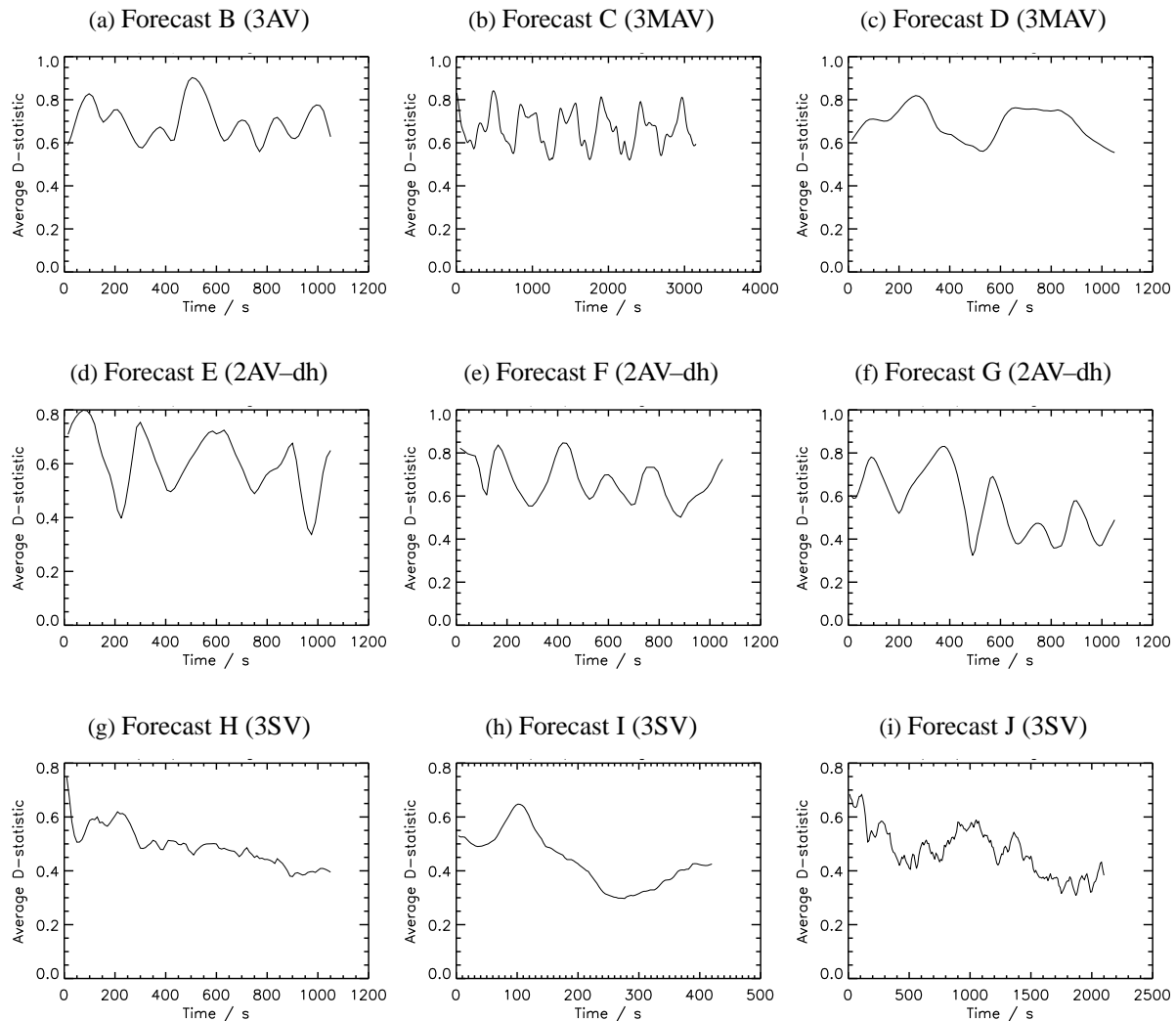
(a) Forecast B (3AV) (b) Forecast C (3MAV) (c) Forecast D (3MAV)

(d) Forecast E (2AV–dh) (e) Forecast F (2AV–dh) (f) Forecast G (2AV–dh)

(g) Forecast H (3SV) (h) Forecast I (3SV) (i) Forecast J (3SV)

Fig. 16. Area-weighted Kolmogorov-Smirnov statistic D as a function of time over the forecast stage. The average is over the vertical slice at $\phi=0.147$ rad. The lower the value of D, the better the climatology is relative to the ensemble as a predictor of the truth.

regimes are more important. We feel that our conclusions about relative predictability are robust, because even though only one forecast was done for each case, three cases were examined in each regime (taking the two quasi-periodic regimes together), with the same conclusions for each. Our results show that it is possible to distinguish between regimes on the basis of their predictability, and this is an important step towards measuring the absolute predictability of the annulus.

We are now in the process of extending this work into the imperfect model scenario, using archival laboratory data to verify the forecasts. We are in the process of developing two data assimilation schemes to do this, the first based on the well-established Met Office analysis correction scheme (Lorenc et al., 1991), and the second based on the more experimental gradient descent filter (Judd, 2003). By combin-

ing these schemes with the method detailed in Sect. 3, this will allow us to make hindcasts of real annulus data, which should provide some insight into the predictability of the rotating annulus in both chaotic and non-chaotic regimes.

In general, we expect real flows to be noticeably more unpredictable than the behaviour examined in this paper. Relaxing the perfect model assumption will also be a useful exercise in validation of the annulus model MORALS; previous work comparing simulated and experimental data has been used to study temperature spectra and velocity profiles (Hignett et al., 1985), flow behaviour with internal heating (Read et al., 1997), heat transport (Read, 2003), and has been used to improve the model integration scheme (Read et al., 2000), but has not compared numerical and experimental time series. Finally, because these techniques are being used in a controlled and reproducible environment, we

re-iterate our proposal mentioned in the introduction, that in the future the annulus could be used as a testbed for assimilation and forecasting techniques currently in operational use, and to test the features and limitations of new techniques before they are made operational.

Edited by: Z. Toth
Reviewed by: three anonymous referees

References

Barnes, J. R.: Midlatitude Disturbances in the Martian Atmosphere: A Second Mars Year, J. Atmos. Sci., 38, 225–234, 1981.

Brier, G. W.: Verification of forecasts expressed in terms of probability, Mon. Weather Rev., 78, 1–3, 1950.

Cai, M., Kalnay, E., and Toth, Z.: Bred Vectors of the Zebiak-Cane Model and Their Potential Application to ENSO Predictions, J. Climate, 16, 40–56, 2003.

Collins, M. and James, I. N.: Regular baroclinic transient waves in a simplified global circulation model of the Martian atmosphere, J. Geophys. Res. – Planet, 100, 14 421–14 432, 1995.

Collins, M., Lewis, S. R., Read, P. L., and Hourdin, F.: Baroclinic Wave Transitions in the Martian Atmosphere, Icarus, 120, 344–357, 1996.

Corazza, M., Kalnay, E., Patil, D. J., Yang, S. C., Morss, R., Cai, M., Szunyogh, I., Hunt, B. R., and Yorke, J.: Use of the breeding technique to estimate the structure of the analysis "errors of the day", Nonlinear Proc. Geoph., 10, 233–243, 2003.

Evans, E., Bhatti, N., Kinney, J., Pann, L., Peña, M., Yang, S.-C., Kalnay, E., and Hansen, J.: RISE Undergraduates find that Regime Changes in Lorenz's Model are Predictable, B. Am. Meteorol. Soc., 85, 520–524, 2004.

Francisco, G. and Muruganandam, P.: Local dimension and finite time prediction in spatiotemporal chaotic systems, Phys. Rev. E, 67, 066 204, 2003.

Früh, W. G. and Read, P. L.: Wave interactions and the transition to chaos of baroclinic waves in a thermally driven rotating annulus, Philos. T. Roy. Soc. A, 355, 101–153, 1997.

Gilmour, I.: Nonlinear model evaluation: ι-shadowing, probabilistic prediction and weather forecasting, D. Phil, thesis, University of Oxford, UK, 1998.

Gilmour, I., Smith, L. A., and Buizza, R.: Linear Regime Duration: Is 24 Hours a Long Time in Synoptic Weather Forecasting?, J. Atmos. Sci., 58, 3525–3539, 2001.

Hide, R.: Some experiments on thermal convection in a rotating liquid, Q. J. Roy. Meteor. Soc., 79, 161, 1953.

Hide, R. and Mason, P. J.: Sloping convection in a rotating fluid, Adv. Phys., 24, 47–100, 1975.

Hignett, P., White, A. A., Carter, R. D., Jackson, W. D. N., and Small, R. M.: A comparison of laboratory measurements and numerical simulations of baroclinic wave flows in a rotating cylindrical annulus, Q. J. Roy. Meteor. Soc., 111, 131–154, 1985.

Houtekamer, P. L. and Derome, J.: Prediction Experiments with Two-Member Ensembles, Mon. Weather Rev., 122, 2179–2191, 1994.

Judd, K.: Nonlinear state estimation, indistinguishable states, and the extended Kalman filter, Physica D, 183, 273–281, 2003.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, 341 pages, 2003.

Kalnay, E., Corazza, M., and Cai, M.: Are bred vectors the same as Lyapunov vectors?, in: Proceedings of a Symposium on Observations, Data Assimilation, and Probabilistic Prediction, part of the 82nd AMS 2002 Annual Meeting, Orlando, Florida, 13-17 January 2002, 173–177, 2002.

Kalnay, E., Peña, M., Yang, S.-C., and Cai, M.: Breeding and predictability in coupled Lorenz models, in: Proceedings of a Seminar held at ECMWF on Predictability of Weather and Climate, Reading, UK, 9-13 September 2002, pp. 29–34, 2003.

Leovy, C. B.: The General Circulation of Mars: Models and Observations, Adv. Geophys., 28, 327–346, 1985.

Lorenc, A. C., Bell, R. S., and Macpherson, B.: The Meteorological Office analysis correction data assimilation scheme, Q. J. Roy. Meteor. Soc., 117, 59–89, 1991.

Lorenz, E. N.: Deterministic nonperiodic flow, J. Atmos. Sci., 20, 130–141, 1963.

Lorenz, E. N.: Climate Predictability, in: The physical bases of climate and climate modelling, Vol. 16 of GARP Publication Series, 132–136, World Meteorological Organisation, 1975.

Massey, F. J.: The Kolmogorov-Smirnov Test for Goodness of Fit, J. Am. Stat. Assoc., 46, 68–78, 1951.

Newman, C. E., Read, P. L., and Lewis, S. R.: Investigating atmospheric predictability on Mars using breeding vectors in a general-circulation model, Q. J. Roy. Meteor. Soc., 130, 2971–2989, 2004.

Patil, D. J., Hunt, B. R., Kalnay, E., Yorke, J. A., and Ott, E.: Local Low Dimensionality of Atmospheric Dynamics, Phys. Rev. Lett., 86, 5878–5881, 2001.

Peña, M. and Kalnay, E.: Separating fast and slow modes in coupled chaotic systems, Nonlinear Proc. Geoph., 11, 319–327, 2004.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T.: Numerical Recipes in Fortran 77, Cambridge University Press, Cambridge, UK, 2nd edn., 1992.

Primo, C., Rodríguez, M. A., López, J. M., and Szendro, I.: Predictability, bred vectors, and generation of ensembles in space-time chaotic systems, Phys. Rev. E, 72, 015 201(R), 2005.

Pu, Z.-X., Kalnay, E., Parrish, D., Wu, W., and Toth, Z.: The Use of Bred Vectors in the NCEP Global 3D Variational Analysis System, Weather Forecast., 12, 689–695, 1997.

Read, P. L.: A combined laboratory and numerical study of heat transport by baroclinic eddies and axisymmetric flows, J. Fluid Mech., 489, 301–323, 2003.

Read, P. L., Bell, M. J., Johnson, D. W., and Small, R. M.: Quasi-periodic and chaotic flow regimes in a thermally driven, rotating fluid annulus, J. Fluid Mech., 238, 599–632, 1992.

Read, P. L., Lewis, S. R., and Hide, R.: Laboratory and numerical studies of baroclinic waves in an internally heated rotating fluid annulus: a case of wave/vortex duality?, J. Fluid Mech., 337,

155–191, 1997.

Read, P. L., Collins, M., Fruh, W.-G., Lewis, S. R., and Lovegrove, A. F.: Wave Interactions and Baroclinic Chaos: A Paradigm for long Timescale Variability in Planetary Atmospheres, Chaos Soliton. Fract., 9, 231–249, 1998.

Read, P. L., Thomas, N. P. J., and Risch, S. H.: 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, 2000.

Smith, L. A. and Gilmour, I.: Accountability and Internal Consistency in Ensemble Formation, in: Proceedings of a Workshop held at ECMWF on Predictability, Reading, UK, 20-22 October 1997, pp. 113–127, 1999.

Toth, Z. and Kalnay, E.: Ensemble Forecasting at NMC: The Generation of Perturbations, B. Am. Meteorol. Soc., 74, 2317–2330, 1993.

Toth, Z. and Kalnay, E.: Ensemble forecasting at NCEP, in: Proceedings of a Seminar held at ECMWF on Predictability, Reading, UK, 4–8 September 1995, vol. 2, 39–60, 1996.

Toth, Z. and Kalnay, E.: Ensemble Forecasting at NCEP and the Breeding Method, Mon. Weather Rev., 125, 3297–3319, 1997.

Toth, Z., Kalnay, E., Tracton, S. M., Wobus, R., and Irwin, J.: A Synoptic Evaluation of the NCEP Ensemble, Weather Forecast., 12, 140–153, 1997.

Tracton, M. S. and Kalnay, E.: Operational Ensemble Prediction at the National Meteorological Center: Practical Aspects, Weather Forecast., 8, 379–398, 1993.

Wei, M., Toth, Z., Wobus, R., and Zhu, Y.: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system, Tellus, 60A, 62–79, 2008.

Yang, S.-C., Cai, M., Kalnay, E., Rienecker, M., Yuan, G., and Toth, Z.: ENSO Bred Vectors in Coupled Ocean-Atmosphere General Circulation Models, J. Climate, 19, 1422–1436, 2006.

Young, R. M. B. and Read, P. L.: Flow transitions resembling bifurcations of the logistic map in in simulations of the baroclinic rotating annulus, Physica D, in press, 2008.